(12) UK Patent Application (19) GB (11) 2 270 419 (13) A

(43) Date of A Publication 09.03.1994

(21) Application No 9320218.2

(22) Date of Filing 10.01.1991

Date Lodged 30.09.1993

(30) Priority Data (31) 463290

(32) 10.01.1990

(62) Derived from Application No. 9213519.3 under Section 15(4) of the Patents Act 1977

(71) Applicant(s)

Microunity Systems Engineering Inc

(Incorporated in USA - California)

843 Channing Avenue, Palo Alto, California 94301, **United States of America**

(72) inventor(s) James A Matthews (51) INT CL5 H01L 23/52

(52) UK CL (Edition M) H1K KGFA K1AA9 K1CA K1FJ K11A3 K11C1A K11C1B K11D K11D3 K4C11 K4C14 K9C1 K9C2 K9D1 K9E K9N2 **K9N3 K9R2**

(56) Documents Cited None

Field of Search UK CL (Edition L) H1K KGF KHAB KHAX INT CL5 HO1L

(74) Agent and/or Address for Service Potts, Kerr & Co 15 Hamilton Square, BIRKENHEAD, Merseyside, L41 6BR, United Kingdom

(54) Interconnect lines for integrated circuits

In a process for fabricating an integrated circuit (IC) which includes a plurality of devices coupled together by a system of metal interconnects, a method of forming said system above the surface of a semiconductor substrate comprising the steps of:

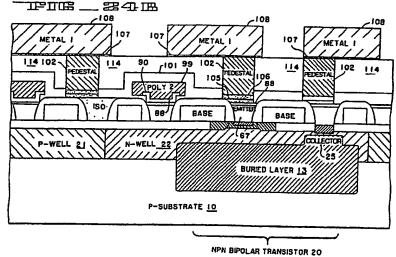
forming a plurality of conductive pedestals (102) on said surface of said substrate, a first portion of said pedestals forming electrical contacts to said devices, said pedestals being formed to a predetermined height such that the top surface of said pedestals is higher than any feature of said substrate;

depositing a first layer of polyimide over said substrate to a thickness which is sufficient to cover said pedestals;

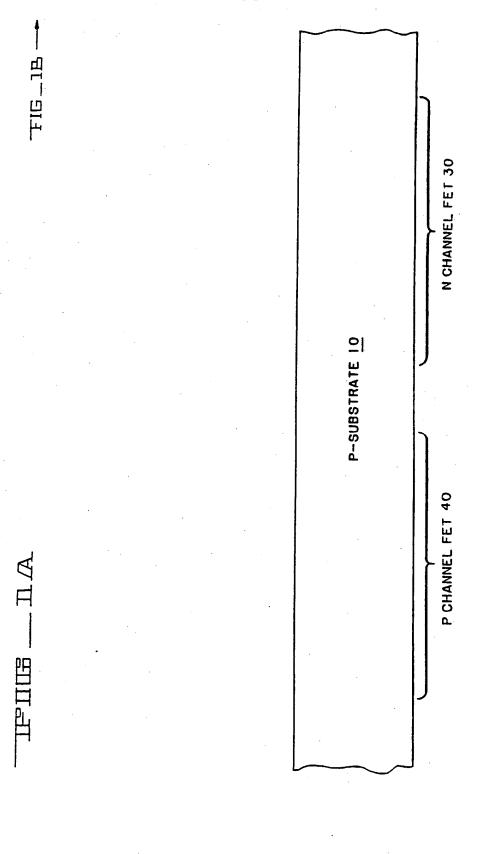
anisotropically etching said first polyimide layer until the upper surface of said first polyimide (c) layer is roughly coplanar with said top surface of said pedestals; and

forming a first set of metal interconnect lines (108) over said first polyimide layer, thereby electrically coupling selected ones of said pedestals.

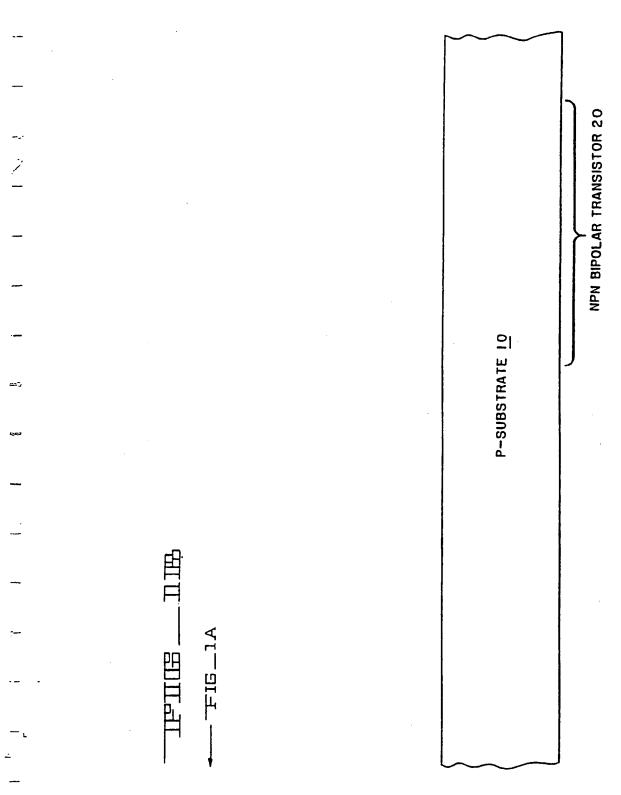
To reduce parasitic capacitance the polyimide layer is etched away to form a system of air-bridge interconnects.

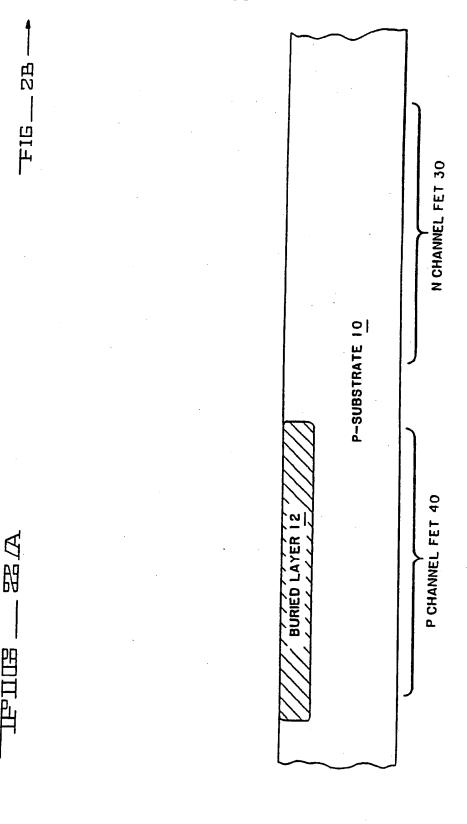


The date of filing shown above is that provisionally accorded to the application in accordance with the provisions of Section 15(4) of the Patents Act 1977 and is subject to ratification or amendment.

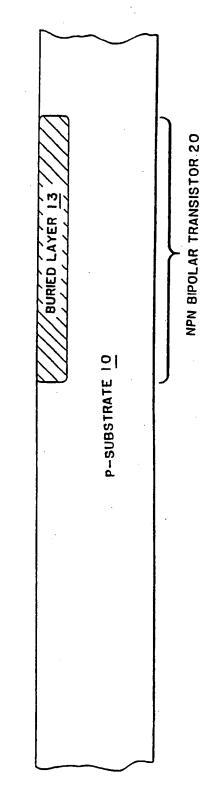


BNSDOCID: <GB 2270419A I :

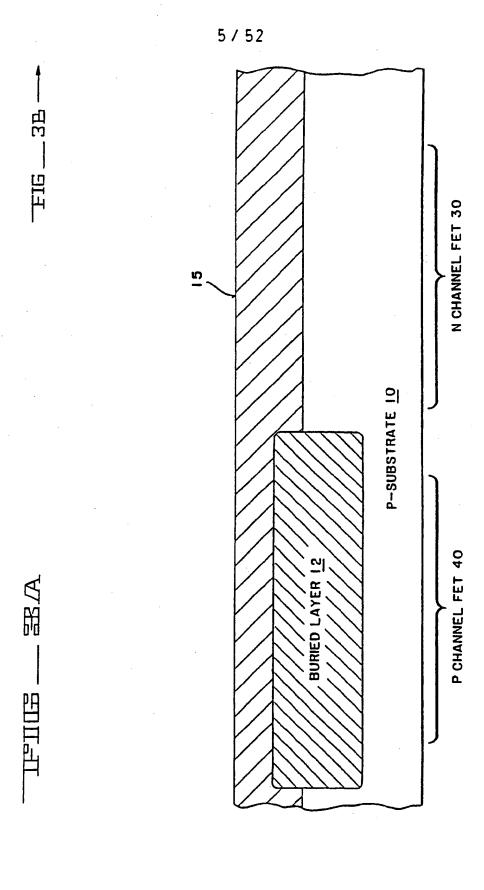




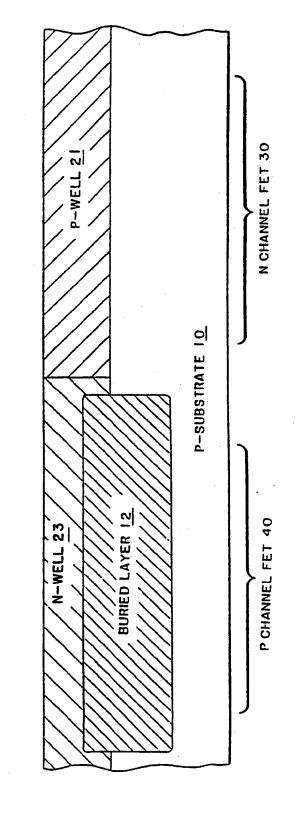
BNSDOCID: <GB___2270419A__I_>



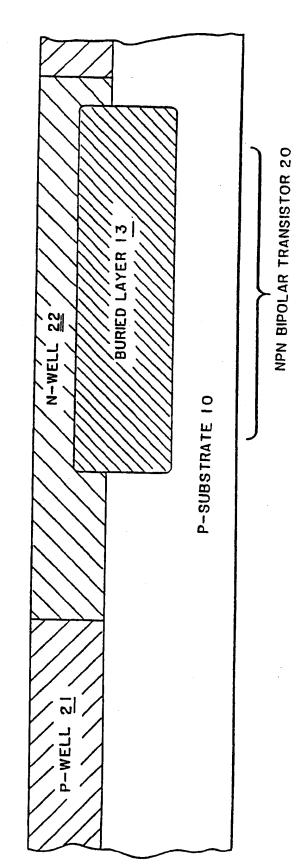
BNSDOCID: <GB___2270419A __I .>



一下工作工品—— 器田



□ — 4A



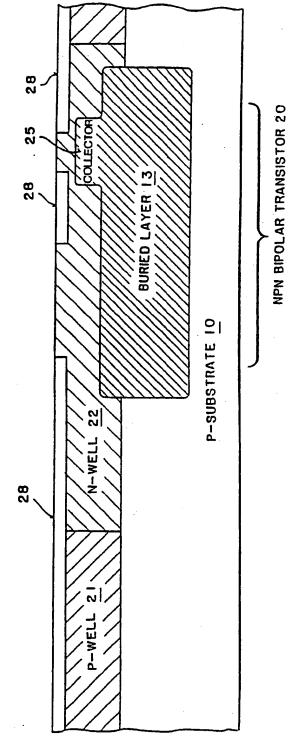
| 正耳 | 一日田

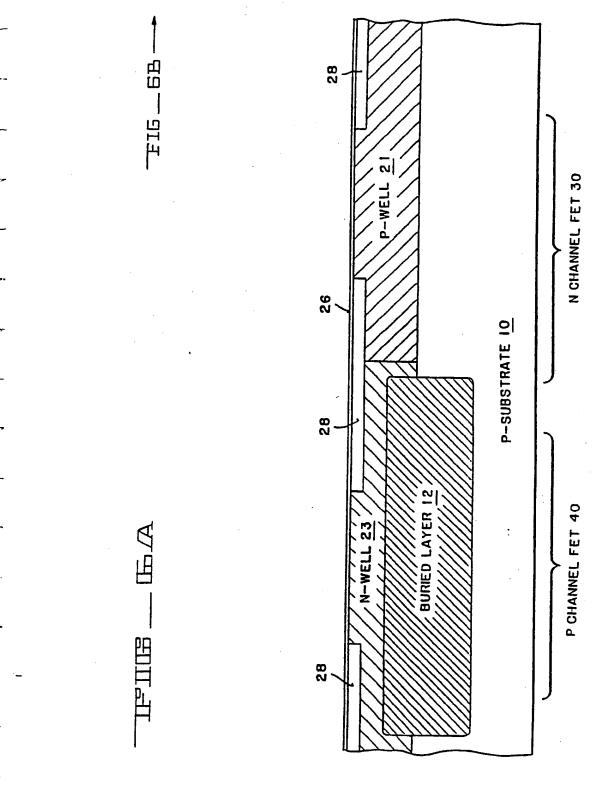
BNSDOCID: <GB___2270419A__I_>

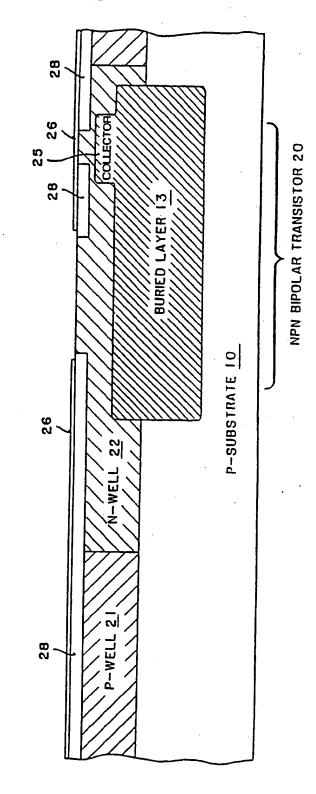
9/52

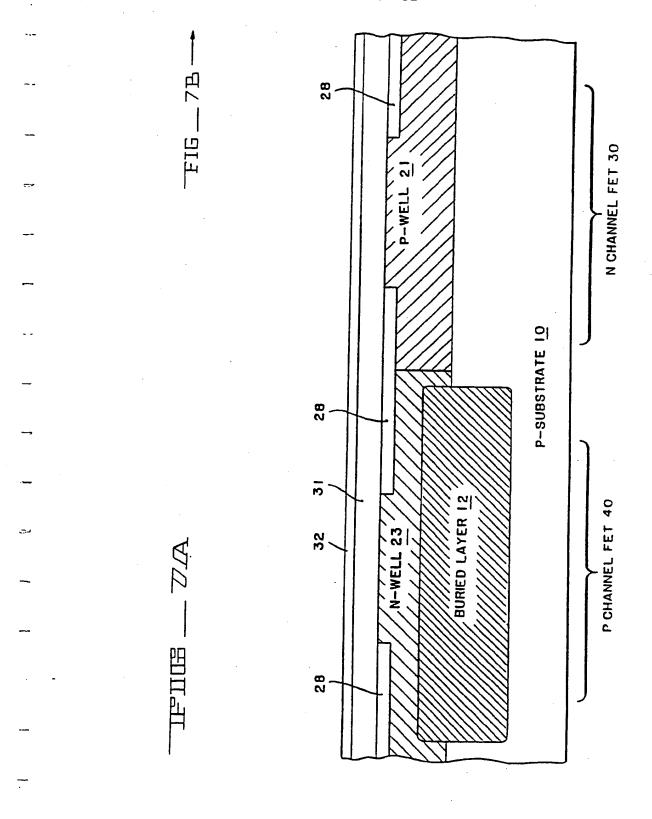
BNSDOCID: <GB___2270419A_1_>

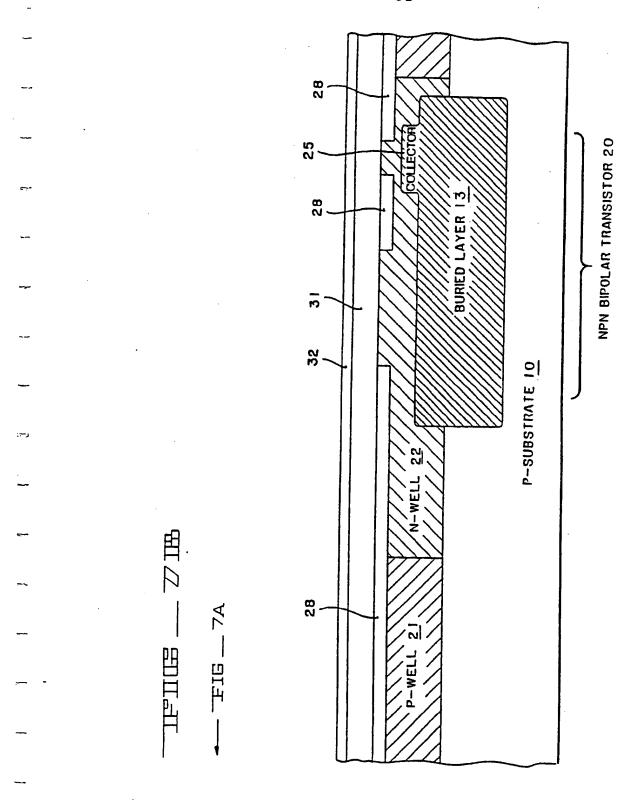
<u>_</u>

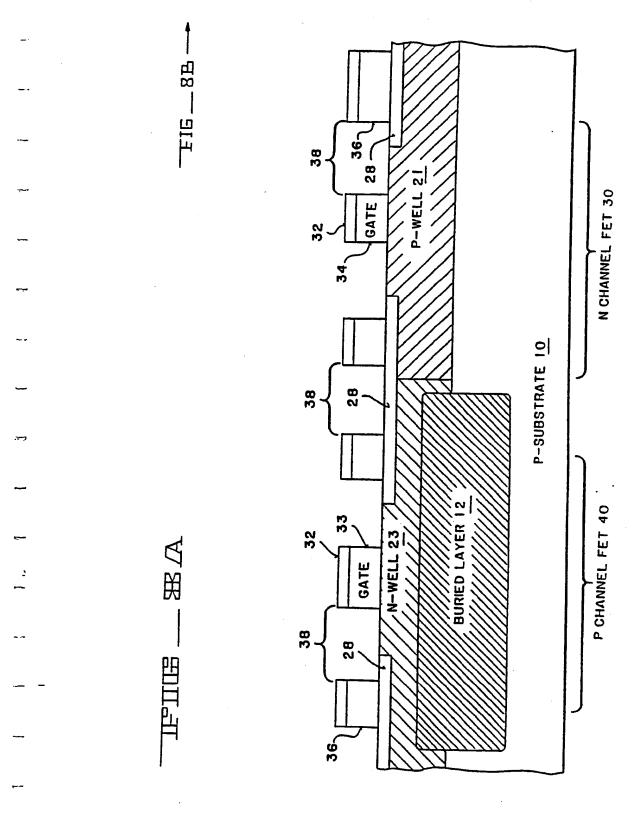


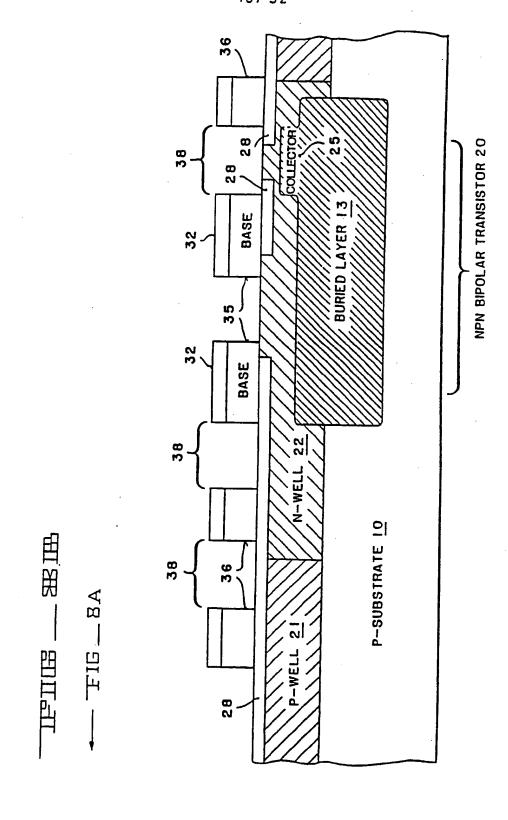




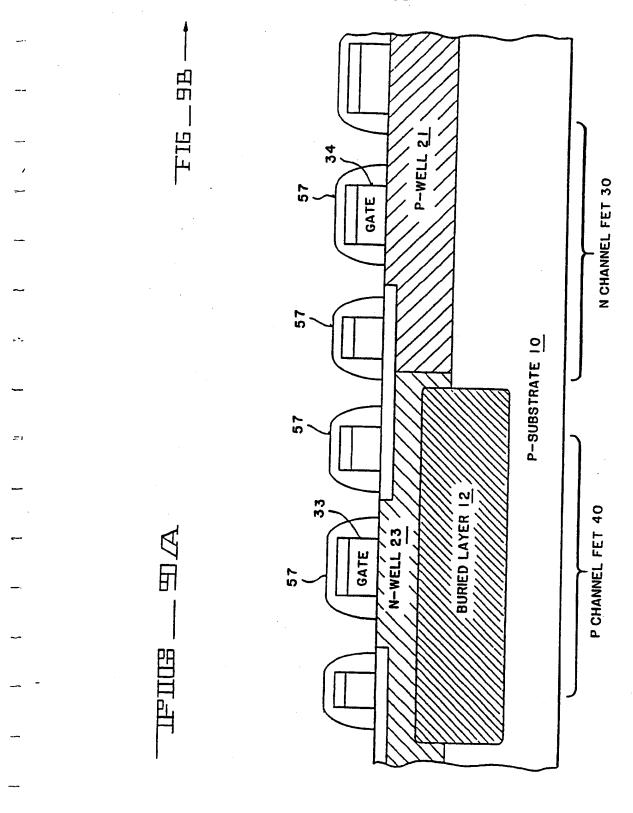




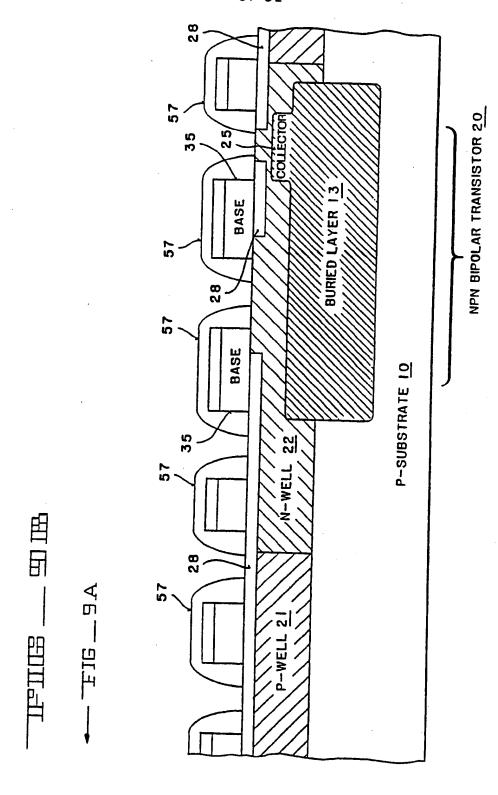




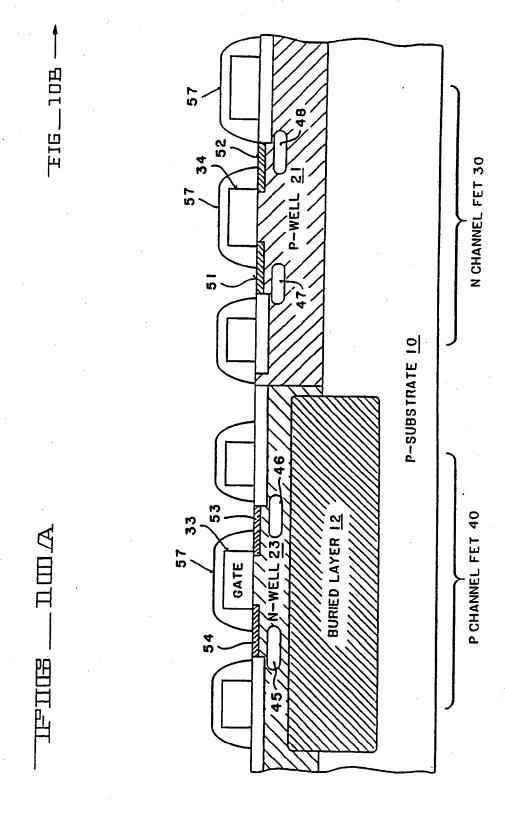
BNSDOCID: <GB___2270419A_I_>



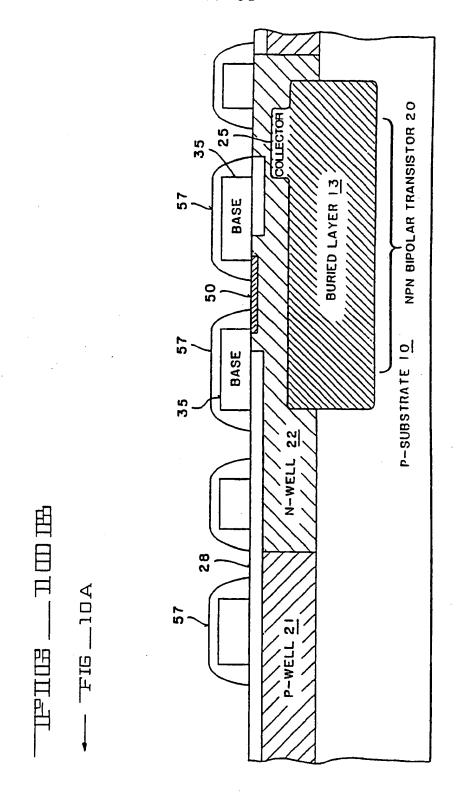
BNSDOCID- - CR 22704194 1 5



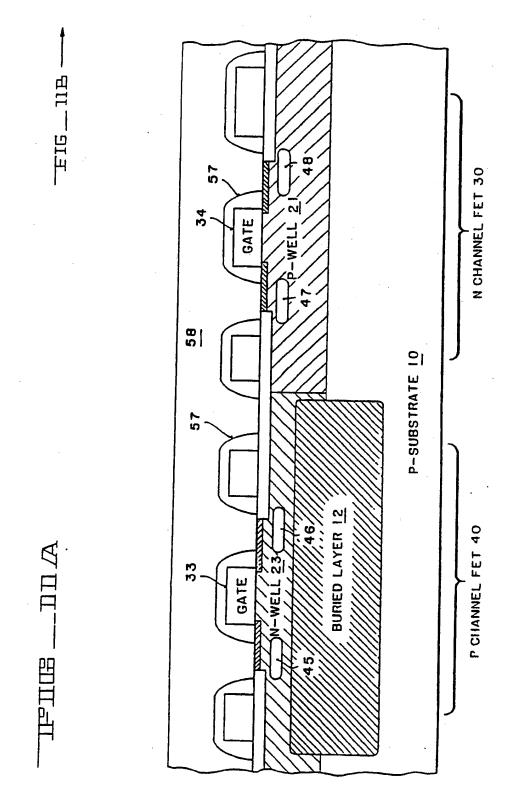
BNSDOCID: <GB 2270419A I >

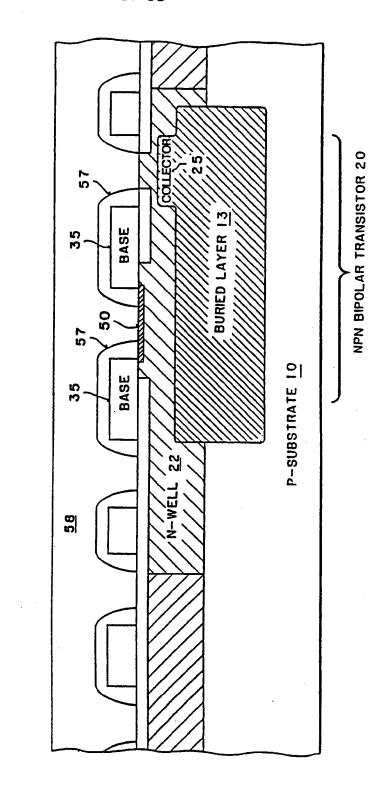


BNSDOCID- ZGB 22704194 I

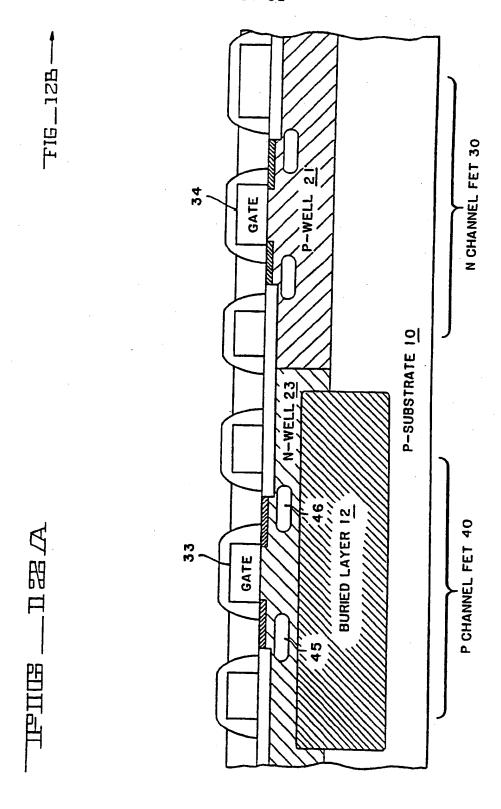


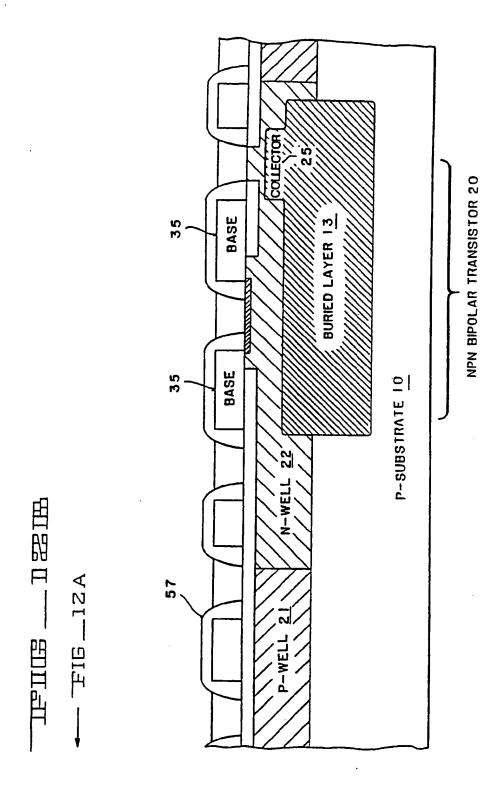
BNSDOCID: <GB 2270419A 1 >



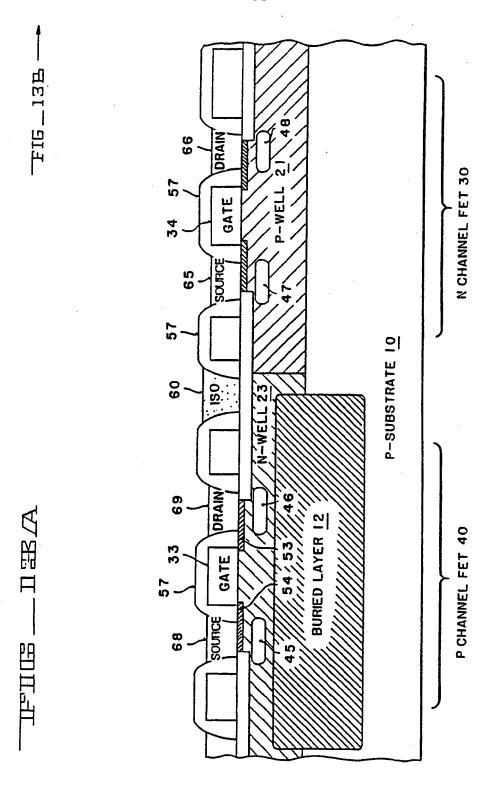


— TE II (B) — — 11 11 (B) → — — — F (F) → 11 P.





BNSDOCID <GB 2270419A | >



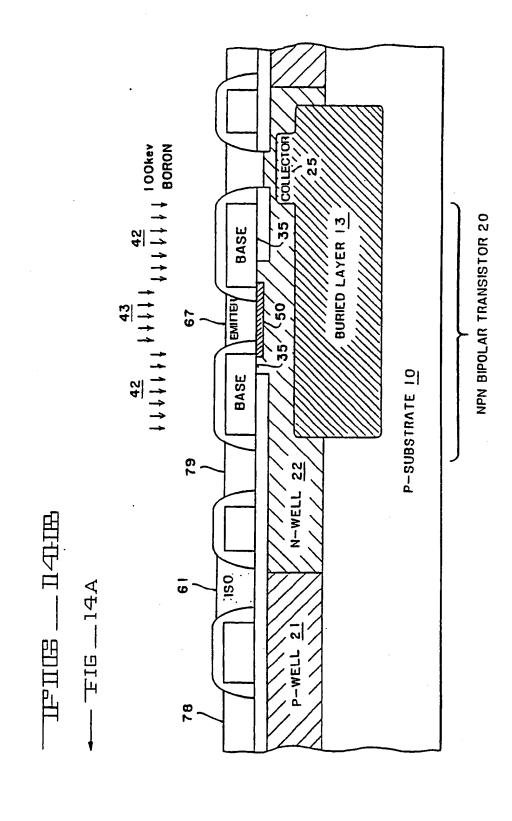
~**-**

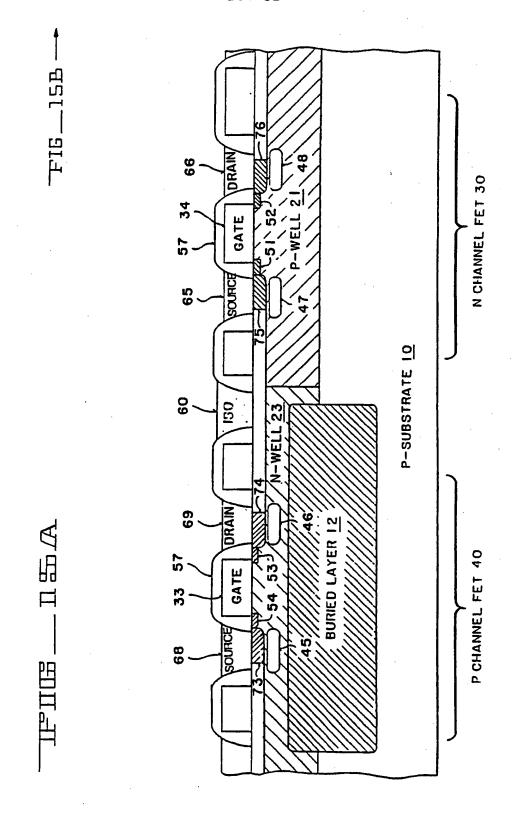
_.

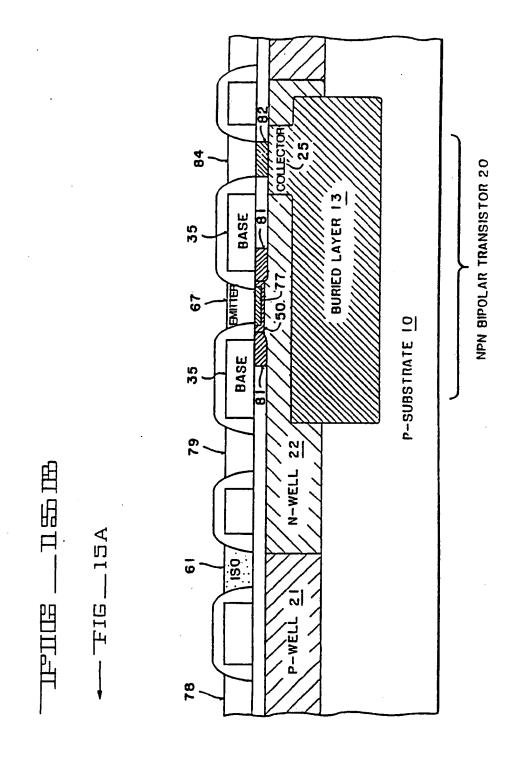
CHEDOCID: +CD 00704404 L-

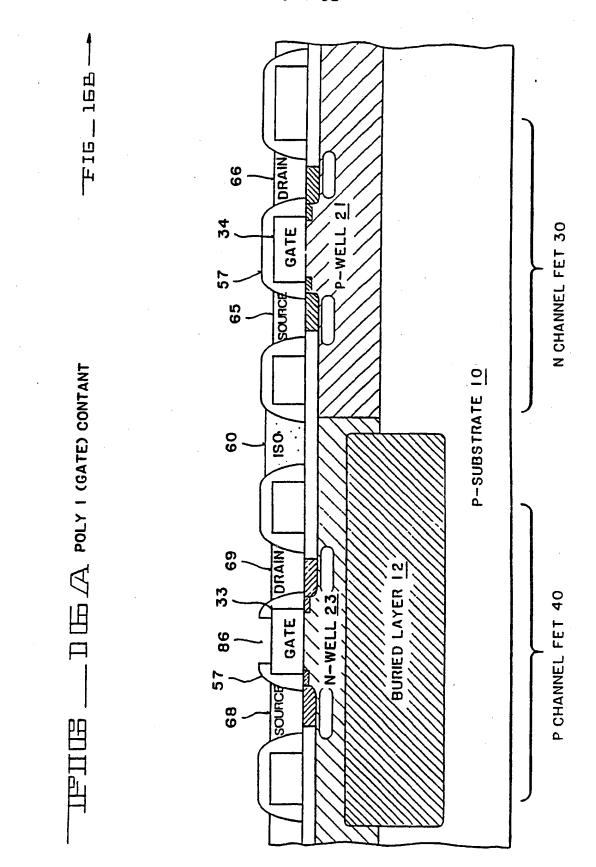
上二山岩 —— 11 部 128 1728 ---- 下16 — 13 A

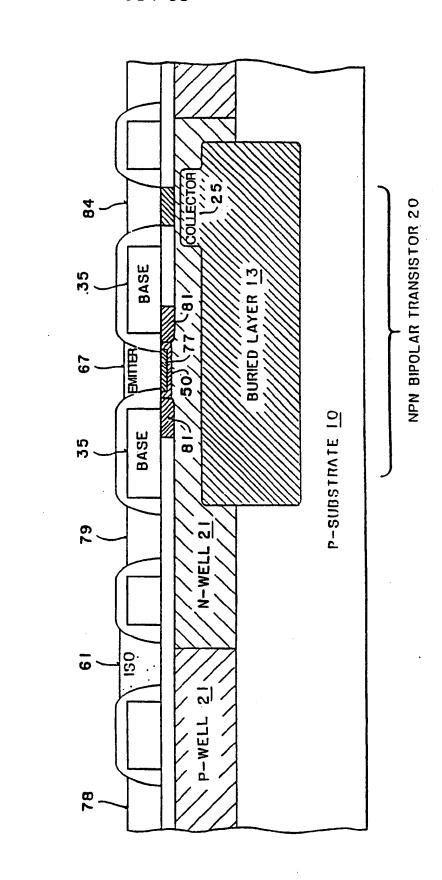
BNSDOCID: <GB 2270419A 1 >







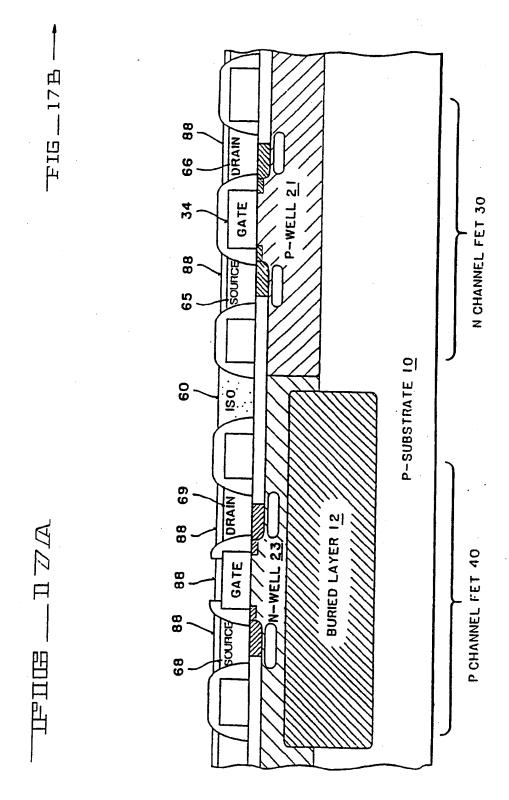




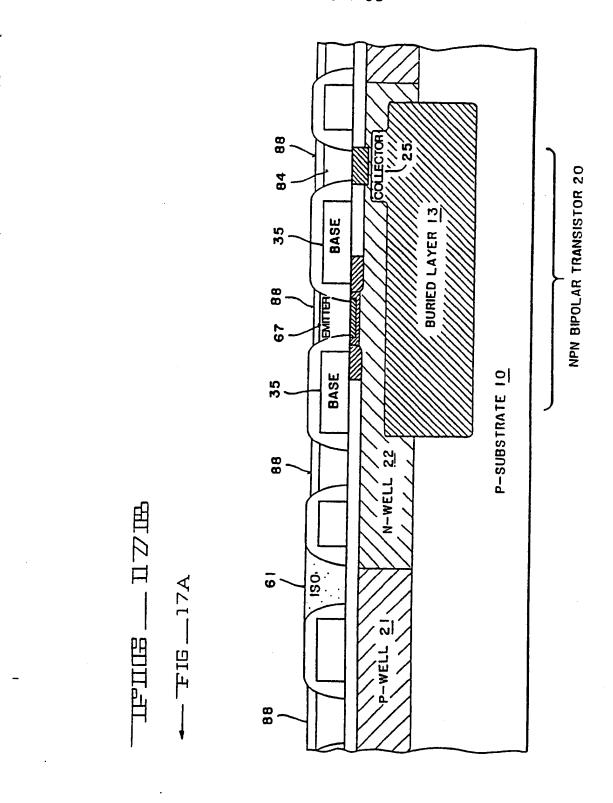
BNSDOCID: <GB 2270419A ... 1, >

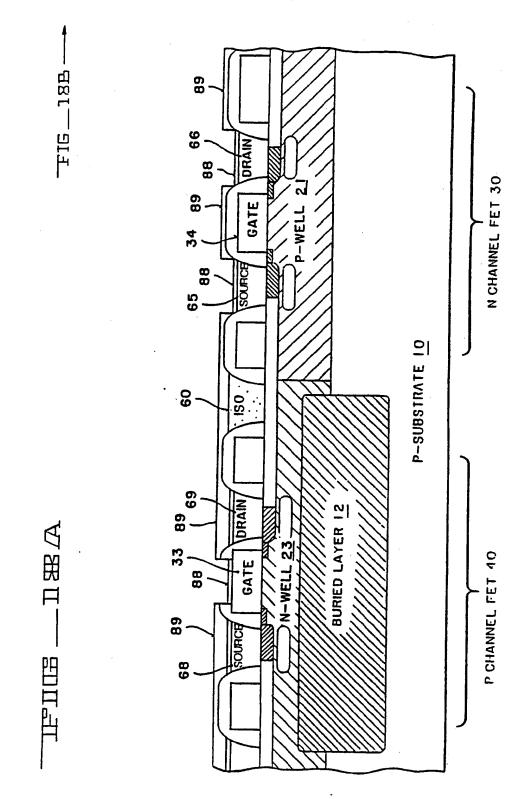
山田田田

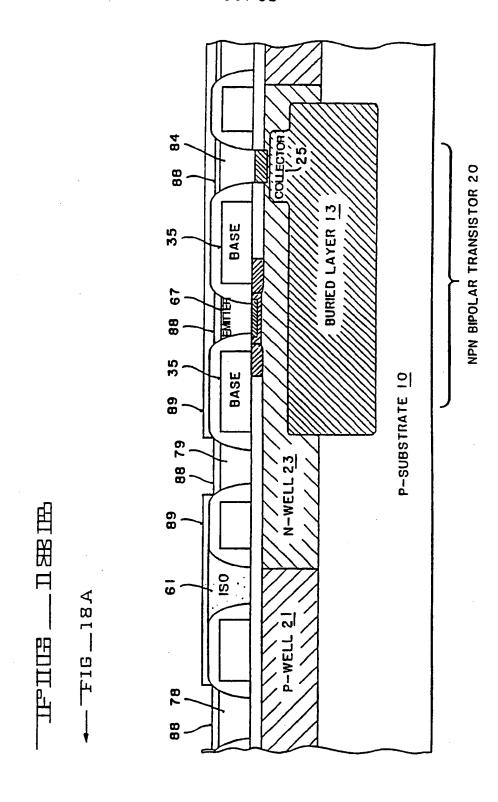
16A



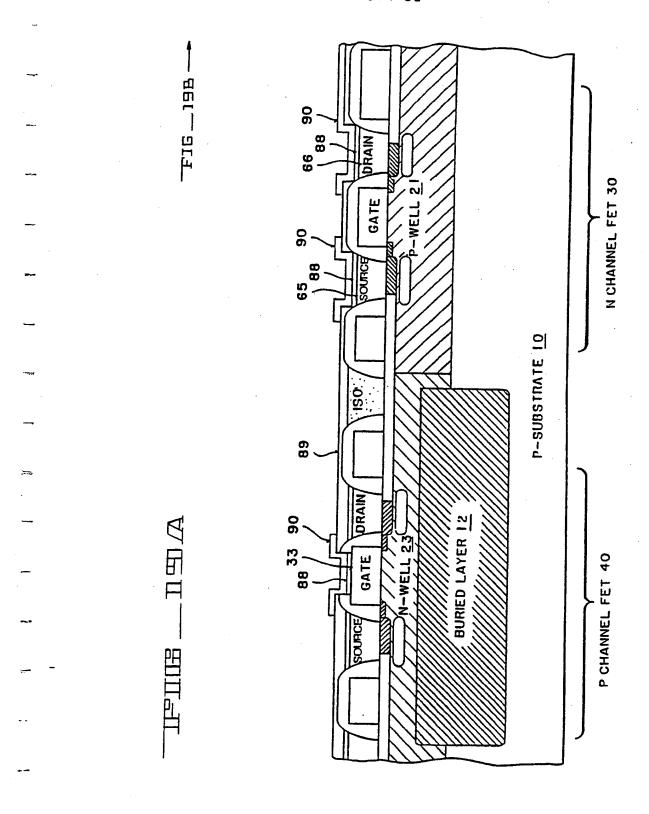
BNSDOCID: <GB __ 2270419A__I_



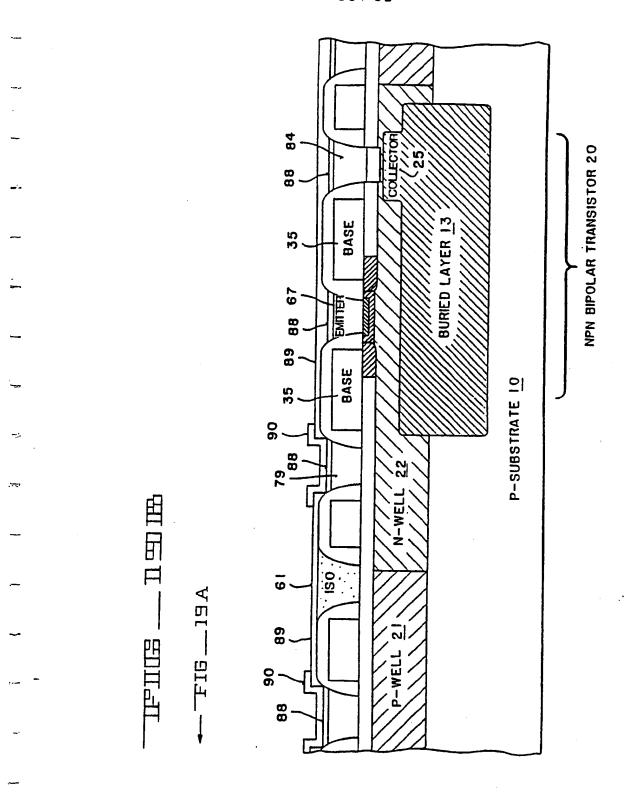


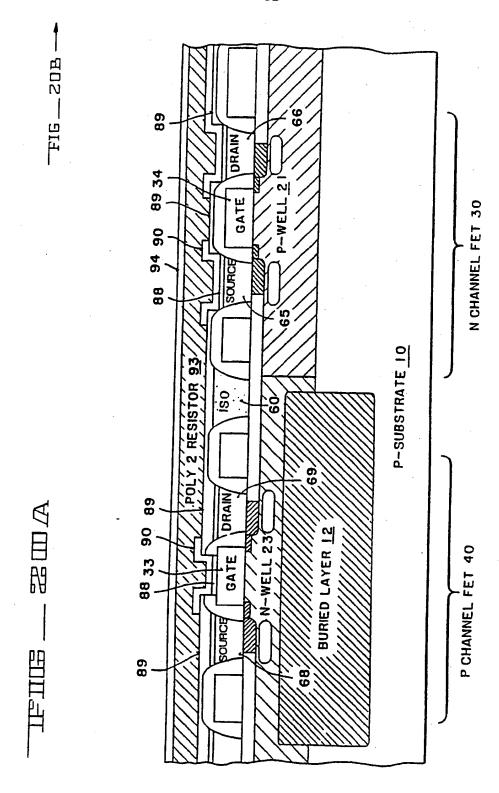


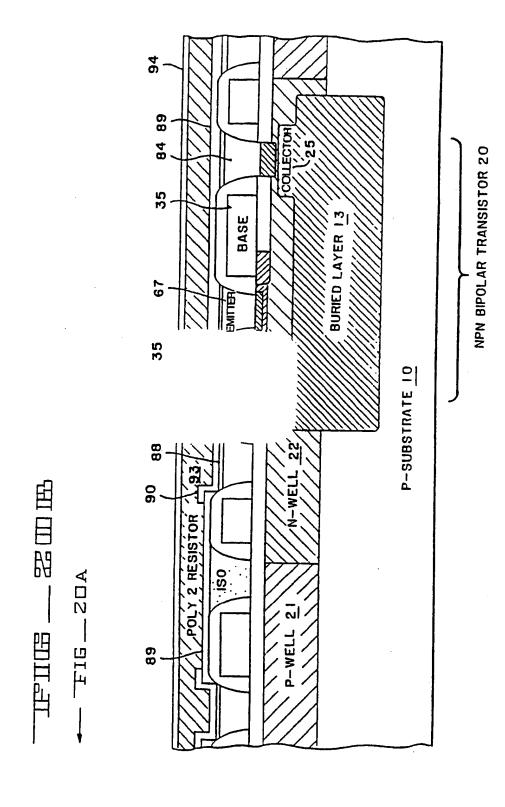
RNSDOCID - CR 22704194 1 3



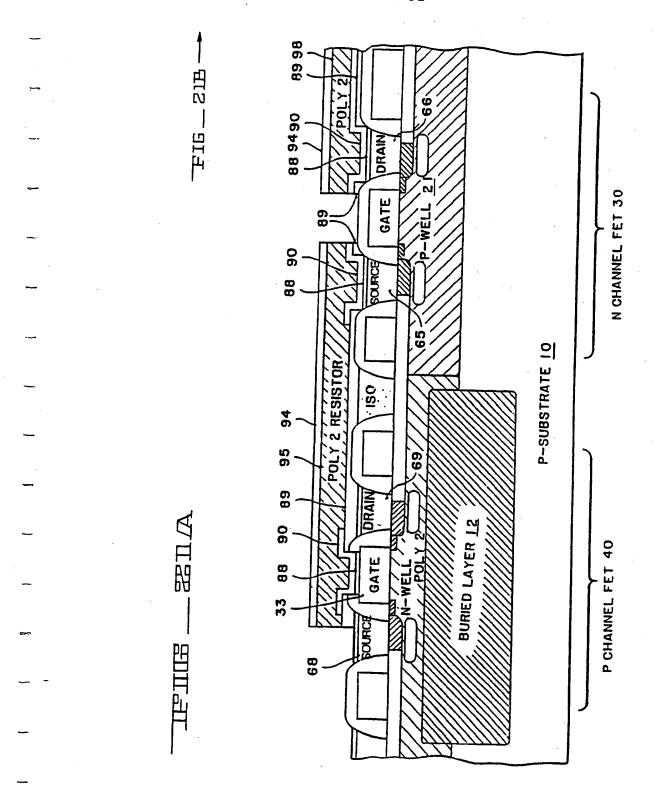
BNSDOCID--GB 22704104 1

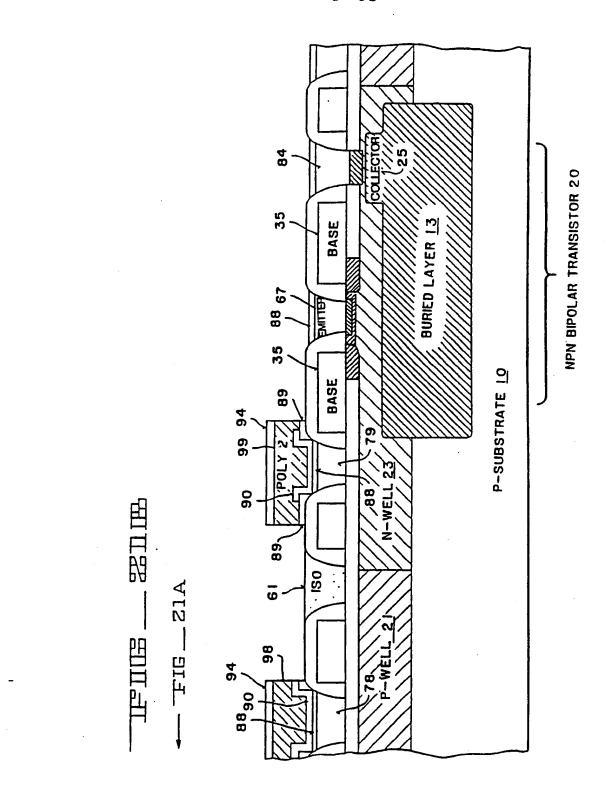


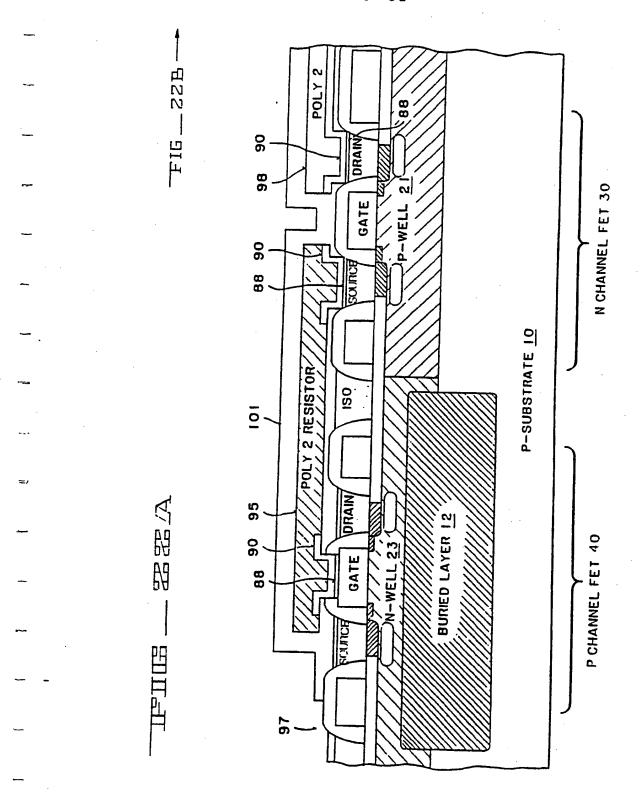


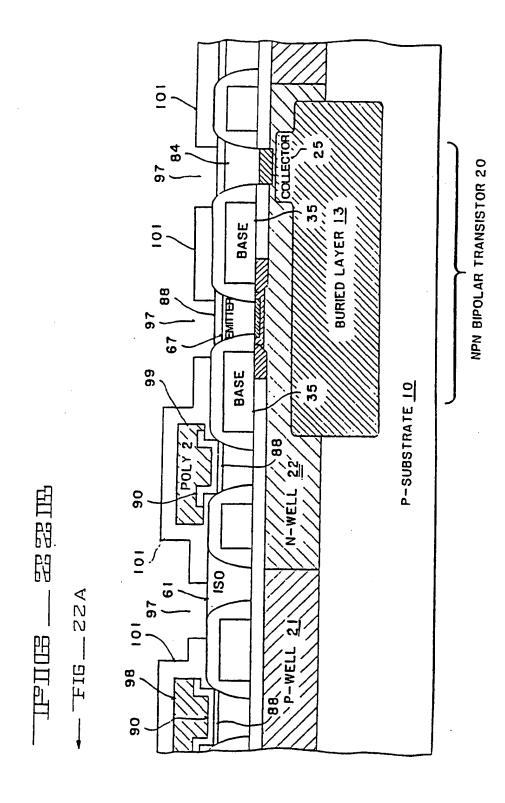


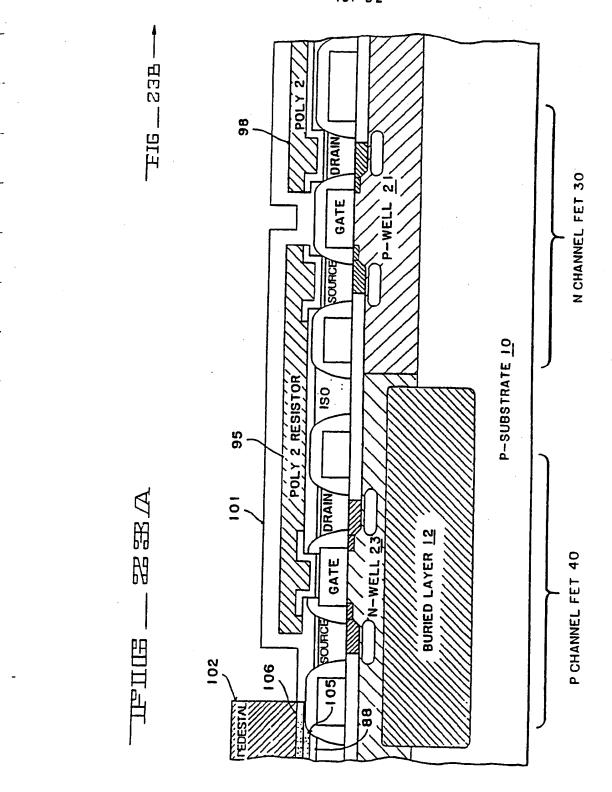
BNSDOCID: <GB___2270419A_ I_>

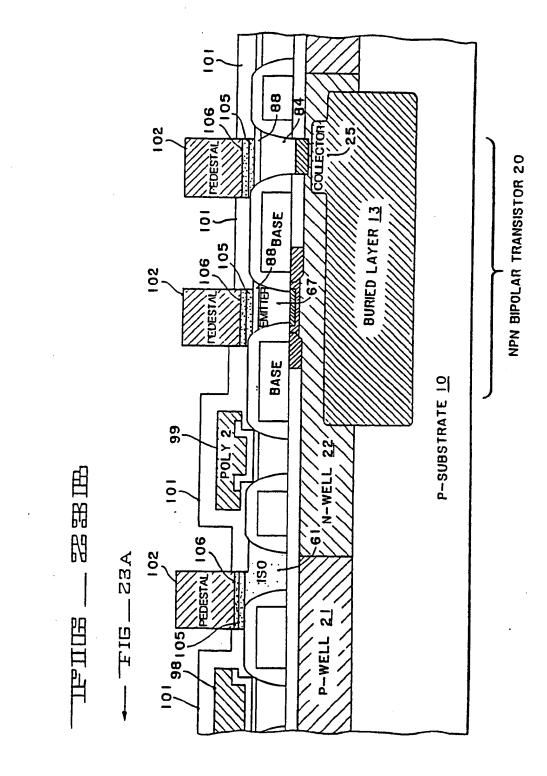


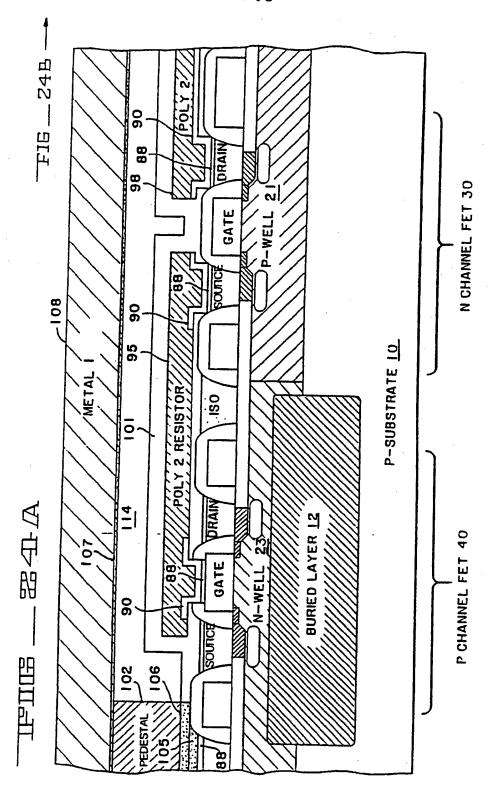


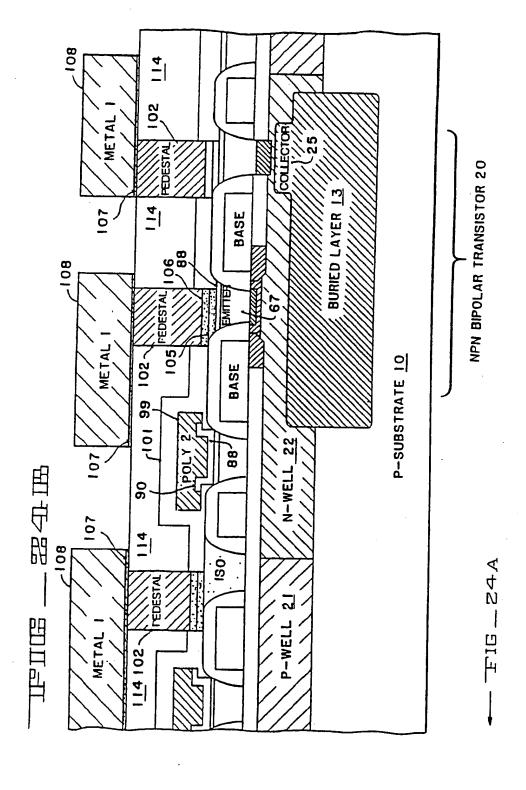


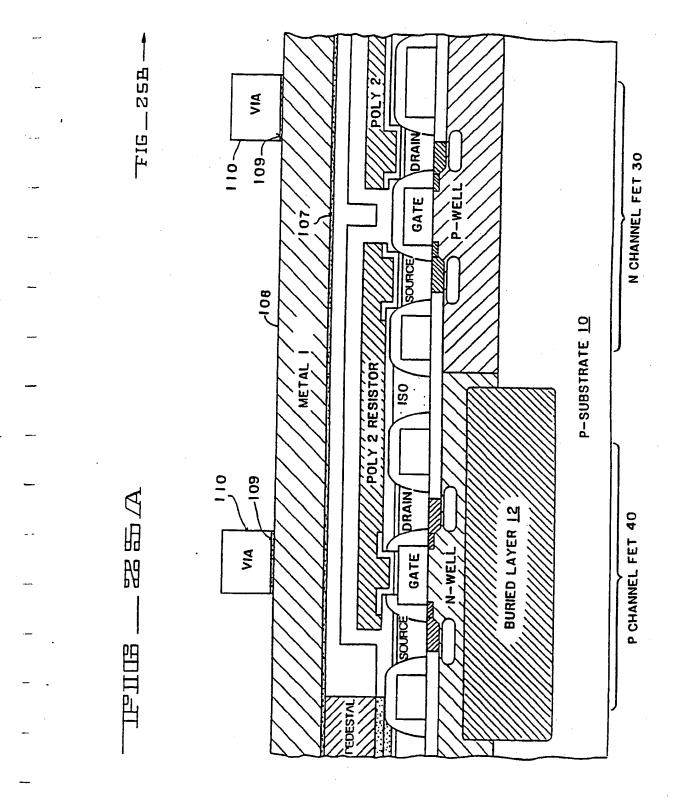


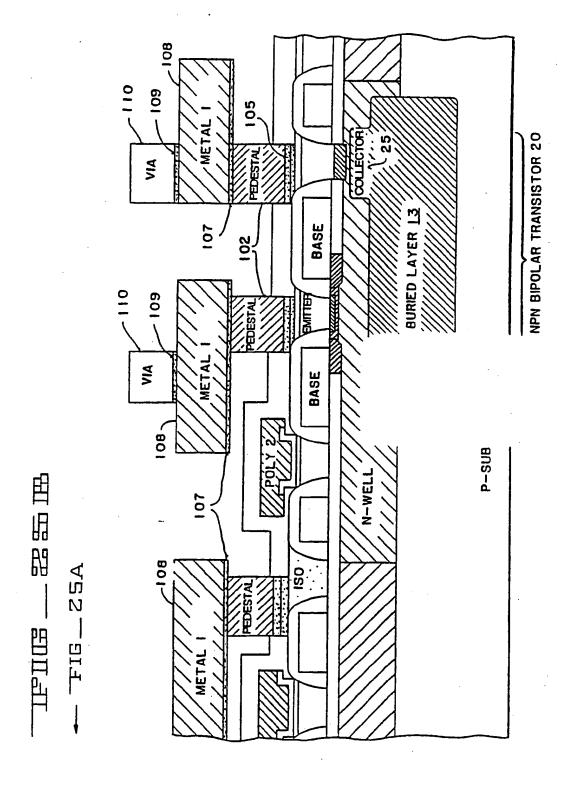


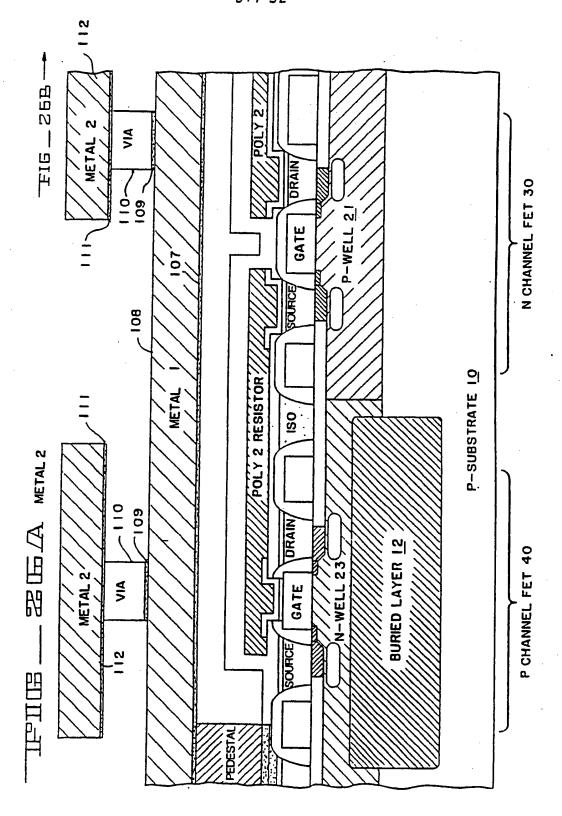


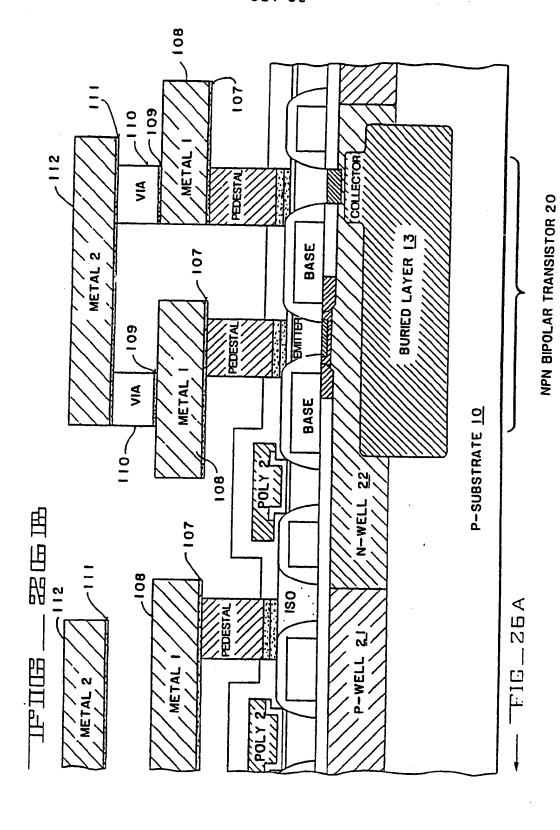












BICMOS PROCESS UTILIZING NOVEL PLANABIZATION TECHNIQUE

FIELD OF THE INVENTION

5

15

20

30

The present invention relates to processes for simultaneously fabricating bipolar and complimentary field-effect transistors in a semiconductor substrate.

O BACKGROUND OF THE INVENTION

In recent years much effort has been directed toward perfecting a method of integrating bipolar and complimentary metal oxide semiconductor (CMOS) technologies on a single wafer. The ability to combine CMOS with bipolar processes in a single ("BiCMOS") process is extremely desirable for high performance circuits. For example, CMOS transistors are inherently low power devices with large noise margins that can achieve a high packing density. Meanwhile, bipolar transistors provide advantages in switching speed and current drive. Bipolar transistors are also characterized by high transconductance which is well suited for driving capacitive loads.

One of the difficulties presented in integrating an MOS device with a bipolar device in the same circuit is that the labrication steps required to form each of the separate devices often differ radically. That is, the steps used to labricate a bipolar device are vastly different from the steps required to labricate a CMOS or an MOS device. As a result, conventional processes are often lengthy and complicated, using a large number of masking operations and numerous thermal cycles.

Another basic limitation to prior an processes has to do with the nature of optical lithography and its affect on the attainable circuit density. In the widely used technique of projection printing, an image of the patterns on the mask is projected onto a resist coated water. Upon the exposure of an Individual chip site the water is moved or stepped on an interferometrically

controlled XY table to the next site and the process is repeated. Using state-of-the-art optics, projection printing (also frequently referred to as direct-step-on-water or step-and-repeat) systems are capable of producing sub-micron resolutions.

However, this high level of resolution does not come without a corresponding trade-off. The trade-off in this case is the limited depth of focus over which the image quality is maintained. For projection printing, the depth of the focus is approximated by the equation

d. f. (depth of focus) = $\lambda/2$ (NA)2

where NA is the numerical aperture of the projection optics and λ is the exposure wavelength. Thus, high resolution (very large numerical aperture) is achieved at the expense of a very shallow depth of focus.

In other words, the ability to print a highly dense circuit layout, having minimal device sizes and structures, is especially dependant on the existence of a highly planar surface on which geometric shapes from a mask may be transferred. Unfortunately, the requirement of a highly planar surface is completely at odds with orthodox methods of forming a semiconductor circuit in a silicon substrate.

In the silicon substrate layer is first subject to exidation. Openings are etched in the exide, and then impurities are introduced or implanted into the substrate. Next, the silicon surface is either reexidized or subjected to depositions of polysiticon. CVD exide, silicon nitride, etc. The result of these successive processing steps (e.g., exidation, etching, implantation, reexidation) are large steps or incongruities running across the surface of the water. Obviously, ever the course of the entire process these steps or incongruities lead to a non-uniform, non-planar water surface. Consequently, the ability to maintain high image quality (due to the depth of focus problems described above) is substantially degraded in prior methods.

One way in which practitioners have attempted to minimize this problem is to planarize the surface of the wafer by reflowing a boron-

5

15

20

phosphosilicate glass across the water surface prior to contact mask. The surface is then aggressively planarized using sophisticated etchants when the metalization steps are reached. Despite these attempts to replanarize the water surface in the back-end processing steps, the tack of planarization in the front-end processing steps (i.e., those steps leading up to the contact mask step) has already taken its toll on the devices. The inability to pattern compact and high-resolution device structures brings about low-density, low-performance circuits. Hence, past semiconductor processes, and particularly BiCMOS processes, have not been able to take full advantage of the high numerical aperture (i.e., resolution) which modem optical equipment can produce due to the depth of focus problem. The critical importance of maintaining planarization throughout each and every processing step is therefore appreciated.

As will be seen, the presently invented BiCMOS process maintains an extremely high level of planarization throughout all of the processing steps using a novel technique known as "waffelization". When combined with a number of additional novel processing features (each of which is believed to be separately inventive in its own right) the disclosed BiCMOS process is capable of producing device dimensions and circuit densities well beyond the limits of the prior art. For example, using the presently invented process it is possible to produce gate widths of 0.5 microns or tess on MOS-type devices and emitter widths of roughly 0.2 microns for bipolar transistors. Consequently, it is contemplated that a 6-transistor memory cell may easily be fabricated within an area of about 3.0 x 4.8 microns -- or a total dimension of approximately 14.4 square microns. This is about the size of a via contact opening in many prior processes.

In addition to achieving high circuit densities and high device performance through planarization of the water surface, the invented BiCMOS process is also characterized by its simplicity, reliability, its self-aligning nature, and the overall design flexibility provided -- both from a circuit design perspective as well as an applications specific viewpoint.

20

Other prior art known to Applicant includes U. S. Patent No. 4,727,046 of Tuntasood et al.; U. S. Patent No. 4,826,783 of Choi et al.; and U. S. Patent No. 4,816,423 of Havemann.

SUMMARY OF THE INVENTION

The present invention disclosed and claimed herein describes a process whereby MOS transistors and bipolar transistors are formed together in the same silicon substrate.

According to the present invention there is provided in a process for fabricating an integrated circuit (IC) which includes a plurality of devices coupled together by a system of metal interconnects, a method of forming said system above the surface of a semiconductor substrate comprising the steps of:

- (a) forming a plurality of conductive pedestals on said surface of said substrate, a first portion of said pedestals forming electrical contacts to said devices, said pedestals being formed to a predetermined height such that the top surface of said pedestals is higher than any feature of said substrate;
- (b) depositing a first layer of polyimide over said substrate to a thickness which is sufficient to cover said pedestals;
- (c) anisotropically etching said first polyimide layer until the upper surface of said first polyimide layer is roughly coplanar with said top surface of said pedestals; and
- (d) forming a first set of metal interconnect lines over said first polyimide layer, thereby electrically coupling selected ones of said pedestals.

BRIEF DESCRIPTION OF THE DRAWINGS

The novel features believed characteristic of the invention are set forth in the appended claims. The invention itself, however, as well as other features and advantages thereof, will be best understood by reference to the detailed description which follows, read in conjunction with the accompanying drawings, wherein:

Figures 1A & 1B are cross-sections of a semiconductor structure

10 illustrating the starting water material and the separate regions for forming

NPN bipolar, n-channel field-effect, and p-channel field-effect transistors.

Figures 2A & 2B are cross-sections following formation of an n+buried layer. . .

15

Figures 3A & 3B are cross-sections following formation of an epitaxial layer.

Figures 4A & 4B are cross-sections after formation of separate p 20 and n-well regions.

Figures 5A & 5B are cross-sections after formation of field oxidation regions.

Figures 6A & 6B are cross-sections after opening an area for the NPN bipolar base region.

Figures 7A & 7B are cross-sections following deposition of the first polysilicon layer.

30

Figures 8A & 8B are cross-sections following gate etch.

Figures 9A & 9B are cross-sections following implantations of n and p-type lightly doped regions, the bipolar base regions, and the well compensation implants.

5 Figures 10A & 10B are cross-sections following oxide side wall spacer etch.

Figures 11A & 11B are cross-sections following planarization poly deposition.

10

Figures 12A & 12B are cross-sections after planarization poly etch back.

Figures 13A & 13B are cross-sections after isolation oxide 15 regions have been formed.

Figures 14A & 14B are cross-sections illustrating source, drain, emitter, implants into the planarized poly regions.

Figures 15A & 15B are cross-sections following source, drain and emitter doping and drive-in.

Figures 16A & 16B are cross-sections following opening of the gate contact regions.

25

Figures 17A & 17B are cross-sections after formation of cobalt silicide contact regions.

Figures 18A & 18B are cross-sections following formation of poly/poly contacts.

į

Figures 19A & 19B are cross-sections following deposition and masking of titanium used for poly 2 silicide.

Figures 20A & 20B are cross-sections after deposition of poly 2 and the nitride top layer.

Figures 21A & 21B are cross-sections after poly 2 mask.

Figures 22A & 22B are cross-sections following masking for 10 pedestal contacts.

Figures 23A & 23B are cross-sections following formation of pedestal contacts.

15 Figures 24A & 24B are cross-sections following metal 1 mask.

Figures 25A & 25B are cross-sections following formation of via pedestals.

Figures 26A & 26B are cross-sections after formation of metal 2 Interconnection regions.

DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

5

10

25

30

A novel BiCMOS process for simultaneously forming bipolar transistors and MOS transistors in the same semiconductor substrate is disclosed. In the following description, numerous specific details are set forth, such as specific conductivity types, dopants, thicknesses, etc., in order to provide a more thorough understanding of the present invention. It will be obvious, however, to one skilled in the art that these specific details are not necessary to practice the present invention. In other instances, other well-known processing steps and methods have not been shown in detail in order to avoid unnecessarily obscuring the present invention.

FORMATION OF THE ACTIVE REGIONS FOR THE BJTs AND MOSFETS

15 Referring to Figures 1A & 1B, a cross-sn- starting semiconductor substrate 10 is shown. Subst I orientation of <100> and has been doped with e impurity to a resistivity of approximately 10 ohm-centil <100> is used because it provides the lowest possible: J. This is 20 standard practice for MOS processing. Optionally, the back side of the wafer may be coated with a polysilicon layer as a means for introducing dislocations into the crystalline structure -- the dislocations acting as impurity trapping sites for heavy metals.

Next, alignment marks are etched into the surface of the wafer using a reactive ion etch (RIE). These alignment marks are in the form of crosses, squares, etc., and are used as an alignment reference for subsequent masking layers. The alignment marks are formed using an alignment mask aligned to the wafer in accordance with standard photolithographic techniques. (e.g., depositing photoresist, masking the surface of the substrate, exposing the masked substrate to ultraviolet light, developing the photoresist, etc.). Considering the high degree of planarization achieved by the invented processes, reference alignment marks provide a simple and

easy method of maintaining mask registration throughout later processing steps.

It should be understood that the alignment mask is an optional masking step in the presently invented process. Alternatively, visual alignment may be made to the implanted buried layer regions. However, since the surface of the wafer will be highly planarized throughout the entire process -- making visual mask registration difficult -- the inclusion of the alignment mask step is preferred. Furthermore, because of the contribution each oxide step has toward defeating planarization, it is desirable to avoid generating a buried layer oxide boundary edge simply for alignment purposes.

With reference to Figures 2A & 2B, buried layer regions 12 and 13 are formed using conventional ion implantation techniques. Antimony ions are implanted at an energy of approximately 180 KeV and at a dose of approximately 1.0 x 10¹⁵ atoms per square centimeter (atoms/cm²). Buried layer 13 forms the n+ collector region for NPN bipolar transistor 20 while buried layer region 12 forms an n+ doped region under p-channel transistor 40.

Following the antimony ion implantation, a p-type epitaxial layer 15 is deposited over the surface of the wafer as shown in Figures 3A & 3B.

Epitaxial layer 15 is approximately 2 microns thick and is doped p-type to a concentration of about 1.0 x 10¹⁷ atoms/cm³. Layer 15 is formed by exposing the wafers to a dichlorosilane (SiH₂Cl₂) gas at approximately 1050° C. The antimony ions in buried layer regions 12 and 13 are subsequently driven (i.e., diffused) downward into p-substrate 10 and upward into epitaxial layer 15 at a temperature of about 1100° C. for approximately five hours in an atmosphere consisting of N₂ + 5% 0₂. The drive-in cycle helps to repair damage done to the substrate surface during ion implantation.

Next, an n-well mask is employed to mask off those regions of epitaxial layer 15 which will subsequently become p-well regions. The

exposed areas of epitaxial layer 15 are then subjected to a phosphorous ion implant at an energy of 180 KeV and a dose of 2.5 X 10¹³ atoms/cm². After this implant, epitaxial layer 15 is divided into separate n-well and p-well regions as shown in Figures 4A & 4B.

In Figures 4A & 4B, n-well region 23 is utilized in the formation of p-channel field-effect transistor (FET) 40. P-well region 21 furnishes the proper conductivity type region for the formation of n-channel FET 30. N-well region 22 is similarly utilized in the formation of NPN bipolar junction "transistor (BJT) 20.

Referring to Figures 5A & 5B, another high-energy implant is employed to form the collector plug region 25, which functions as a low resistance vertical contact down to the underlying n+ buried layer region 13. As mentioned earlier, buried layer region 13 eventually forms the collector of NPN transistor 20. After appropriate masking, phosphorous ions are implanted at an energy of 180 KeV and a dose of 1.0 x 10¹⁵ atoms/cm². The collector drive and n-well drive are performed simultaneously at 1100° C. for approximately six hours in an atmosphere comprising N₂ + 5% O₂. The N₂ + 5% O₂ gas helps to establish an equilibrium between nitridation and oxidation of the surface of the water. This prevents excessive nitridation of

Following the collector plug and n-well drive cycles, a thin (approximately 250Å) pad silicon dioxide layer is grown across the entire surface of the water. This pad oxide is grown in a furnace at 900° Celsius (C) for fifteen minutes using an oxygen gas flow. A standard silicon nitride deposition using gas phase, low pressure chemical vapor deposition (LPCVD) techniques follows growth of the pad oxide. This silicon nitride layer acts as a mask layer during formation of the field oxide regions. Growing a pad oxide beneath the nitride layer markedly reduces stress induced dislocations in the underlying silicon substrate. (Note that the pad oxide is not shown in the Figures because of the relative thinness of that layer.)

The growth of the field oxide (FOX) regions are performed using the well-known processing method known as recess oxidation (ROX). In the recess oxidation method a thick field oxide layer is first grown thermally in the patterned FOX regions. Due to the fact that thermally grown oxide takes up more volume than does single crystalline silicon, there is a substantial volume of silicon dioxide which grows above the planar surface of the substrate. This elevated portion of the silicon dioxide is subsequently etched isotropically to produce a recessed surface across the wafer. After "etching the wafer surface is again oxidized, resulting in a nearly planar surface. Active regions for the devices are defined by the FOX regions at this point in the process.

Normally, field exidation produces what is commonly referred to as birds'-head or birds'-beak features. These formations consist of an exide bump (birds'-head) along with a lateral growth of exide (birds'-beak) at the silicon nitride boundaries. The length of the birds'-beak is proportional to the thickness of the field exide layer. Typically, lateral exide penetration is between one-half and one times the field exide thickness. The birds'-beak phenomena sets a fundamental limit on the pitch of the active region; that is, how close active layers can be placed adjacent to one another.

In the presently invented process, field oxide regions are defined using a high resolution active layer mask which opens selected areas in the silicon nitride layer. The nitride is etched anisotropically using a reactive ion etch (RIE). A first field oxide layer is then grown in these openings at 900° C in steam to form a 2000Å oxide film. This first field oxide film is then dipped off in a 10:1 solution of hydrogen fluoride (HF) 48% and water. A second field oxide layer is then grown at 900° C in steam to form an approximately 1000-2000Å field oxide layer 28 as shown in Figures 5A & 5B.

Because the field oxide thickness in the invented process is relatively thin when compared to prior an methods, the penetration of the birds'-beak into the active layer regions is markedly reduced. By way of example, active layer spacing is on the order of 1.0 microns. Previously, attempts to reduce

10

15

20

25

the thickness of the field oxide region were avoided since a thin field oxide generally corresponds to a high coupling capacitance between the substrate and the overlying metal interconnects. The reason why a relatively thin field oxide is tolerated in the presently invented process is because the interconnection layers are not in contact with, nor in close proximity to, the substrate surface. This aspect of the present invention will be described later.

In addition, because field plate isolation is used in the presently invented BiCMOS process, the surface potential is easily controlled. Also, use of field plate isolation and thin field oxide contributes greatly to the radiation hardness of the circuit. This is of critical importance in certain military applications. Figures 5A & 5B show field oxidation regions 28 after the silicon nitride masking layer has been removed.

Yet another advantage of using a thin field oxide is that by choosing a silicon nitride mask layer thickness carefully, the field threshold adjustment implant can penetrate through the field oxide after it is grown, as well as the silicon nitride layer. In this way, a single implant may be utilized to provide correct V_T for the transistors while simultaneously providing the proper field threshold level.

When used in this manner, the silicon nitride acts as a sort of decellerating or breaking layer -- setting a shallow depth for the implant dose in the transistor active regions. Optionally, the silicon nitride layer can be removed prior to the field implant, possibly resulting in significantly higher channel mobilities due to the deeper implant. Higher channel mobility translates into better transistor performance due to the corresponding reduction in impurity scattering.

The ability to perform transistor threshold and field implants simultaneously also provides great flexibility in selecting device types. For example, if boron is implanted in an n-channel device an enhancement-mode n-channel MOSFET having a threshold of approximately 0.5 volts is produced. On the other hand, if this implant is left out of the active regions

10

15

20

25

30

ر زپ

(by appropriate masking) then a natural device (i.e., approximately 0 volt V_T) results. Furthermore, if the p-type enhancement-mode device implant is directed into the n-channel transistors, a -0.5 volt V_T depletion mode device is created.

Moreover, a single masking/implant process step can be used to change the characteristics of selected field-effect devices from enhancement-mode, to depletion-mode, or to zero threshold type devices. Because of the symmetry between n-channel and p-channel devices each individual transistor is available in all three types. This results in a substantial savings in masking steps and allows interesting device type combinations. By way of example, it might be desired to produce a NAND gate which uses an enhancement-mode lower device and a zero threshold upper device.

Ion implantation is a most valuable tool for controlling threshold

voltage. Very precise quantities and purities can be introduced making it

possible to maintain extremely close control of VT. For instance, where

boron is implanted through a gate oxide for a p-channel device, the

negatively charged boron acceptors serve to reduce the effects of the

positive surface state charge. As a result, VT becomes less negative.

Similarly, a shallow boron implant into the p-type substrate of an n-channel

transistor can make VT positive -- as required for an enhancement device.

The capability, as described above, of allowing enhancement, depletion,
and zero threshold type devices to be incorporated on the same chip using a
single implant therefore provides many advantages to the integrated circuit

In the currently preferred embodiment, the n-type V_T and field implant mask aligns to the active layer region. A boron ion implantation follows at an energy of 100 KeV and a dose of approximately 5 x 10¹² atoms/cm². Following this implant, a p-type V_T and field implant is performed. The masking layer is aligned to the active layer and a phosphorous ions are implanted at 180 KeV with a dose of approximately 1 x 10¹³ atoms/cm².

Following the threshold implants, the field nitride regions are reactive ion etched until removed, then the underlying pad oxide is dipped off in a 100:1 solution of water and HF.

FORMATION OF THE GATE DIELECTRIC

5

10

At this point in the process we are ready to form the gate oxide for the n-channel and p-channel FETs. After the pad oxide etch there can exist residual nitride films around the birds'-beak regions. These residual films can suppress subsequent oxidation resulting in a thinner gate oxide in these areas. To insure that this residual nitride is completely removed a first gate oxide layer of approximately 200Å is grown in a furnace at 900° C in an atmosphere of 87% argon, 10% 02 and 3% HCl. Immediately afterwards this first gate oxide is dipped off in a 100:1 solution of HF. This extra growth and etching step insures complete removal of residual nitride films. The second gate oxide (i.e., the actual gate oxide used in the devices) is also grown in a furnace at 900° C using the same gas as was used in growth of the first gate oxide. This second gate oxide layer is carefully grown to a thickness of approximately 100Å.

After the gate oxide has been formed, selected regions must be removed to create buried contacts down to the substrate. One example is in the case of bipolar NPN transistor 20 where a buried contact is needed for electrical connection down to the extrinsic base region. Traditionally, buried contacts are formed by first spinning a tayer of photoresist over the gate oxide, developing the photoresist, etching the gate oxide in the contact areas, then depositing polysilicon. This technique has been widely used in NMOS transistor processes.

Conversely, because the presently invented BiCMOS process employs a gate oxide which is extremely thin (approximately 100Å), it is undesirable to expose this thin oxide to reactive chemicals. Therefore, in the invented process after the growth of the gate oxide layer is completed, a first polysilicon deposition is performed in a turnace at 580° C. This produces a

30

<u>.</u>...

thin (approximately 500Å) layer of amorphous polysilicon which covers all of the gate oxide previously grown. This thin layer of polysilicon acts to cap the very thin gate oxide prior to the formation of the buried contacts.

It is significant that this first protective polysilicon layer is amorphous in structure. By its nature, amorphous polysilicon is highly smooth and contains no embedded grain structure. In contrast, ordinary polysilicon (grown at a higher temperature) is generally much rougher and more granular than amorphous polysilicon. When ordinary polysilicon is reactive ion etched down to the single crystalline silicon substrate the grain structure of the polysilicon can become imprinted into the substrate surface. This causes considerable damage to the underlying crystal lattice and can diminish device performance, particularly in bipolar junction transistors.

Figures 6A & 6B show cross-sections in which the protective polysilicon layer 26 is opened in base contact region 27 using conventional masking techniques. The base contact regions of layer 26 are first etched using a reactive ion etch and then the underlying gate oxide is dipped off in a 100:1 solution of water and HF. (Note that in Figures 6A & 6B the underlying gate oxide layer is not shown due to its relative thinness when compared with other layers). The opening over region 27 is utilized to later form buried base contacts for BJT 20.

An important feature of the invented process is that both n-type and p-type buried contacts are available. This is in contrast with standard MOS processes where all of the polysilicon is typically doped n-type so that buried contacts to p-channel FETs or p-type BJT base regions are generally not available. If an attempt were made to form buried contacts to a p-type substrate using n-type polysilicon the n-type dopant from the polysilicon would diffuse down and form a PN junction with the underlying p-type material. In the presently invented process, p-type polysilicon is used in p-type channel FETs and an n-type polysilicon is utilized on n-channel FETs. Hence, as will be described in more detail below, symmetric buried contacts are available.

15

20

25

FORMATION OF FIRST LEVEL POLYSILICON MEMBERS

After the buried contact regions have been opened, a second, much thicker, layer of polysilicon is deposited over the water. In the preferred embodiment, the second polysilicon deposition occurs in a furnace at 580° C. This produces an amorphous layer which is approximately 2000Å thick. Again, it is important that the second polysilicon deposition be amorphous to maintain an extremely smooth and planar surface.

In Figures 7A & 7B, the amorphous second polysilicon layer is shown as layer 31. Note that polysilicon layer 31 completely covers field exidation regions 28 as well as base contact region 27. It is understood, of course, that there is no gate exide present in base region 27 underlying poly layer 31. As described above, this exide was removed in a previous processing step. But gate exide does underlie poly layer 31 over the channel regions of n-channel FET 30 and p-channel FET 40.

Following the deposition of poly layer 31 a high temperature TEOS (tetra-ethyl-ortho-silicon) oxide 32 is deposited over the surface of poly layer 31. Reoxidation of poly layer 31 is avoided to prevent recrystalization of poly layer 31. A reoxidation cycle might also destroy the amorphous nature of layer 31 and form asperities therein.

Note also that in Figures 7A & 7B, poly layer 31 is Illustrated as a uniformly planar surface for ease of representation. Actually, there exists a physical step over each of the buried contact regions. The height of the step being equal to the thickness of the protective poly layer 26 (see Figures 6A & 6B) plus the thickness of the gate oxide. The sum of the gate oxide and poly layers, and therefore the height of the step, is about 500Å. (For purposes of discussion, the surface of the wafer is considered planar whenever any step or feature is 1000Å or less in height.)

Referring now to Figures 8A & 8B, next in the sequence of steps is

the polysilicon gate etch. To perform the gate etch an extremely high
resolution masking layer is aligned back to the active layer of the devices. A

10

15

20

5

25

reactive ion etch of oxide layer 32 is followed by a second reactive ion etch which anisotropically etches the underlying polysilicon layer. By using a highly selective etchant the anisotropic etch of the polysilicon stops on the thin gate oxide.

Recall that where emitters are to be formed there is no existing gate oxide to stop on. Over etching in these regions is potentially hazardous to the underlying crystal lattice. However, if amorphous silicon is used as described above, no grain pattern is impressed into the substrate. Thus the electrical properties of the emitter junction are maintained. Because of the use of amorphous polysilicon, over etching in the emitter regions does not have a deleterious effect on device performance.

Following the poly gate etch, a slight exidation of the exposed sidewall polysilicon is performed. Approximately 1000Å of exide is grown on the poly in the silicon in a furnace at 900° C in a steam atmosphere.

15 Oxidizing the poly gate sidewalls helps to marginally reduce the gate width while fully insulating each of the polysilicon members. It also creates a minute birds'-beak at the silicon interface which reduces the electric field density near the polysilicon edges and corners. Oxidation of the sidewalls also densifies the top exide layer 32.

At this point in the process, gate 33 of p-channel FET 40 has been defined along with gate 34 of n-channel FET 30 and the extrinsic base contacts 35 of NPN bipolar transistor 20. Additionally, polysilicon members 36 have also been defined along the surface of the wafer. The function of polysilicon members 36 will be explained in more detail later.

Note that in Figures 8A & 8B, uniform spaces 38 have been defined between each polysilicon member. In the currently preferred embodiment, spacings 38 are each about 0.5 microns in width. Maintaining a controlled and accurate polysilicon spacing 38 across the entirety of the wafer is of crucial importance to the goal of planarization. When viewed from the wafer's top, these uniform polysilicon spacings 38 create a "waffle-like" pattern across the surface of the wafer. Hence, the origin of the term

"waffelization" to denote the technique for achieving complete planarization of the semiconductor surface.

Using a controlled poly space 38 (all polysilicon spaces being equal in dimension, i.e., "waffelized") makes possible the invented planarization scheme. This scheme is based on the isotropic (i.e., uniform in all directions) deposition of a material — in this case polysilicon — followed by an unmasked anisotropic (i.e., vertical) etch. Once spaces 38 have been etched out of the polysilicon tayer and a sidewall oxide formed, an isotropic deposition of amorphous polysilicon is used to fill in spaces (i.e., the gaps between the previously patterned polysilicon lines). Since the spaces 38 are uniform across the wafer, each is filled at the same time and at an identical rate.

Deposition continues until a totally planar surface is produced.

Generally, this requires a polysilicon deposition to a thickness which is
approximately twice the thickness of members 36. The anisotropic etch of
this second polysilicon layer reduces the thickness until the second poly
layer is roughly co-planar with poly members 36. At this stage,
members 36 are electrically isolated from the second polysilicon layer. The
second polysilicon layer is now confined to spaces 38 and comprises a
plurality of second poly members 58 (see Figures 12A & 12B). To
summarize, the effect of these processing steps is to fill the gaps between
the poly lines, resulting in a planar surface. Details of these processing
steps, with reference to the pertinent figures, will follow.

spacings 38 to a fixed dimension places no constraint on the width of polysilicon members 33, 34, 35, etc. Thus, it is possible to have base contacts, gates, or polysilicon interconnect members with various widths. First polysilicon can take the form of a gate, an interconnect, or simply a dummy polysilicon member such as members 36 shown in Figures 8A & 8B. In addition, polysilicon members 36 may be employed either as an interconnect layer or simply to establish uniform spaces 38 across the

surface of the wafer. As previously mentioned, defining equally-wide spaces 38 is the key concept to the invented planarization-by-waffelization scheme.

Following the etch of the polysilicon surface which defines members 33-36, oxide sidewalls are formed on each of the polysilicon members. A high temperature TEOS oxide is deposited at 700° C to form a 2000Å thick oxide layer over the wafer. This oxide is then reactive ion etched until the source drain regions and emitter regions appear in the silicon surface. A slight reoxidation is then performed to protect the exposed silicon surface.

Figures 9A & 9B show each of the polysilicon members 33-36 being insulated by silicon dioxide regions 57 following reactive ion etching to create openings in spaces 38 down to the substrate. Notice that a slight slope is etched on each of the sidewall areas of oxide 57 between adjacent first polysilicon members. This slope helps to insure a uniform deposition of polysilicon in spaces 38.

Given that the chief constraint for achieving planarization at this stage in the process is maintaining uniform spaces of polysilicon, a great variety of polysilicon patterns may be realized. For instance, the surface of the water may primarily consist of large (i.e., very wide) spaces 38 with small islands or strips of polysilicon separated throughout. Alternatively, the surface of the water may be very dense with polysilicon and employ only the minimum (e.g., 0.5 microns) space 38 permitted by the process design rules. Thus, a great deal of flexibility is inherent in the topological patterning of the poly 1 material.

FORMATION OF THE SOURCE/DRAIN REGIONS FOR FETS AND BASE/EMITTER REGIONS FOR BJTs

Preferably, after the sidewall oxide regions 57 have been defined,

formation of the source/drain regions for FETs 30 and 40 and base/emitter regions for BJT transistor 20 may proceed. Referring to Figures 10A &10B, p-type lightly-doped drain/source (P-LDD) regions 54 and 53 are first

10

15

20

formed just below the surface of the substrate. The same mask used for the implantation of regions 54 and 53 is also used in forming the intrinsic base region for BJT 20 and the well compensation regions 45 and 46 for p-channel FET 40.

The P-LDD + BJT base + well compensation mask aligns to the gate mask. Once openings in the photoresist are formed, a boron ion implantation at an energy of 30 KeV and a dose of 1.0 x 1014 atoms/cm² produces lightly-doped regions 54 and 53 and intrinsic base region 50. The purpose of the lightly-doped implant is to enhance the reliability of each of the FET devices. If heavily doped n+ or p+ regions are allowed near the gate oxide an extremely high electric field can result. The presence of a high electric field would enhance undesirable hot-electron effects in those transistors. Lightly-doping the source/drain regions near the substrate surface (i.e., adjacent to the gate) substantially reduces the electric field strength. As a result, device degradation due to hot-electron generation is suppressed. (In the preferred embodiment, the junction depth of regions 53, 54 and 50 is approximately 1500Å.)

Note that the boron implant is made directly through the side wall portion of oxide layer 57 which surrounds gate member 33. The blocking effect that oxide 57 has on the implant helps to keep high concentrations of boron impurities away from the channel region. This also helps to reduce the vulnerability of the device to hol-electron effects.

Alternatively, the P-LDD implant can be performed prior to forming the sidewall spacer oxide regions 57. However, if the implant is performed before the oxidation then the boron atoms tend to diffuse laterally under the gate. This has the undesirable effect of increasing overlap capacitance between the gate and the source/drain regions.

Using the same masking step, a second boron ion implantation may be performed. In the preferred embodiment, this second implant is at an energy of 100 KeV and a dose of 1.0×10^{13} atoms/cm². This high energy implant produces deep well compensation regions 45 and 46 in p-channel

20

25

FET 40. (If the same mask is used a well compensation region is also formed in BJT 20. This well compensation implant region is not shown in Figures 10A & 10B since it is not critical to ordinary device performance. Of course, the base of BJT 20 may be protected by a second mask tayer prior to the second boron ion implant to avoid forming a well compensation region in this device.)

Because n-well region 23 is heavily-doped, there is a correspondingly high junction capacitance C_i associated with FET 40. While it is desirable to have heavy doping in the deep channel region to reduce short channel effects of the device, it is equally desirable to reduce C_i . Including additional well compensation implant regions 45 and 46 significantly reduces (by a factor of approximately three) the FET junction capacitance C_i resulting in a higher speed, lower power circuit. Recognize that this well compensation implant costs no additional masks and is self-aligned with the source/drain regions.

Next, N-LDD regions 51 and 52, and well compensation regions 47 and 48 are formed using a single mask step as described above for the P-LDD implants. The N-LDD + well compensation mask aligns to gate 34. A phosphorous ion implant follows at 60 KeV and a dose of 20 1 x 10¹⁴ atoms/cm² to form regions 51 and 52. This is followed by another phosphorous ion implant at 180 KeV with a dose of 1.0 x 10¹³ atoms/cm² to form well compensation regions 47 and 48. As mentioned, lightly-doped regions 51 and 52 help to reduce hot-electron effects in n-channel FET 30. At the same time, well compensation 25 regions 47 and 48 reduce junction area capacitance.

With reference to Figures 11A & 11B, after the N-LDD and well compensation implants have been performed, source, drain, emitter (SDE) polysilicon layer 58 is deposited isotropically over the surface of the wafer in a furnace at 580° C. This is the polysilicon which fills all of the spaces 38 across the wafer. Once again, SDE poly 58 is an amorphous poly layer and is deposited to a thickness of approximately 10,000Å.

15

5

10

20

25

30

Because SDE poly 58 is deposited isotropically, an absolutely uniform deposition takes place on all sides of oxide 57 and above silicon substrate. Uniform deposition rates as well as uniform spacing between polysilicon members insures that previously vacant spaces 38 are filled isotropically. In other words, if the spaces 38 between polysilicon members were nonuniformal or varied in width, then certain regions (narrower spaces) would fill quicker resulting in a nonplanar surface.

After a sufficient thickness of SDE poly 58 is deposited to insure planarization, it is anisotropically (vertically) etched back. This reactive ion etch of SDE poly 58 continues until the tops of the gates are cleared of polysilicon as is shown in Figures 12A & 12B. At this stage of the process, SDE poly 58 fills all of the spaces 38 between polysilicon members 33-36. Eventually, SDE poly members 58 will be used to produce the sources, drains, and emitters of the transistors in addition to forming an interconnect layer for the completed circuit. As is clearly seen in Figures 12A & 12B, following etch back of SDE poly layer 58 the surface of the wafer is entirely replanarized.

Unfortunately at this point, all of the devices are electrically shorted together. Logically then, the next processing step involves electrically isolating each of the devices 20, 30 and 40. To isolate the devices, a LPCVD nitride layer is first deposited over the entire surface of the wafer. Selected areas of the nitride layer are opened using a conventional mask step and portions of the SDE polysilicon are etched down to the underlying field oxide. Nitride protects the tops of all other devices. The etch of the SDE polysilicon is both anisotropic and isotropic so as to guarantee clearing all of the polysilicon out of the cavities.

A high temperature TEOS oxide is then deposited at 700° C to a thickness of approximately 10,000Å. This TEOS oxide is deposited isotropically and fills in the uniform spaces similarly to the SDE poly deposition -- again utilizing waffelization to achieve planarization. The TEOS oxide forms the isolation regions 60 and 61 as illustrated in

Figures 13A & 13B. Recognize that isolation region 61 isolates bipolar transistor 20 from n-channel FET 30 while isolation region 60 isolates p-channel FET 40 from n-channel FET 30 after the oxide is RIE etched to clear the oxide above the nitride. The underlying nitride layer is then etched off leaving the structure shown in Figures 13A & 13B.

The next masking step opens the source, drain and gate regions of n-channel FET 30. All three regions are then implanted with a phosphorous ion implant at 180 KeV with a dose of approximately 3.0 x 10¹⁵ atoms/cm². This phosphorous ion implant is depicted in Figures 14A & 14B by the arrows 41. Phosphorous is deliberately chosen as the n+ type dopant because of its high diffusivity, i.e., its ability to rapidly distribute atoms throughout the polysilicon members. Other SDE polysilicon members -- such as members 78 and 79 which could be used as an interconnect layer -- may also be doped n-type using implant 41.

Once n-channel FET 30 has been implanted with phosphorous, another mask is applied to the water to open the area over p-channel FET 40. This p+ implant mask also opens up the base areas of BJT 20 while simultaneously protecting the emitters. Boron, having a relatively high diffusivity, is implanted at 100 KeV and a dose of 3.0 x 10¹⁵ atoms/cm². This p+ implant, shown in Figures 14A & 14B by arrows 42, dopes polysilicon members 68, 33, 69, and 35 (forming the source, gate and drain polysilicon members of p-FET 40 and polysilicon base member 35 of

An emitter implant follows the p+ implant and is performed in the same manner as described above. The emitter implant mask exposes the emitter polysilicon member 67 of BJT 20. An arsenic ion implant is then done at 180 KeV with a dose of 3.0 x 10¹⁵ atoms/cm². Of course, the energy of the arsenic implant may be adjusted downward. If necessary, to avoid implanting arsenic ions into the intrinsic base region 50. The emitter implant is shown in Figures 14A & 14B by arrows 43.

BJT 20, respectively) heavily p-type.

Note that in addition to opening up emitter regions the emitter implant

15

20

25

mask also may be used to open up areas over those polysilicon members which could become interconnection layers (e.g., members 78 and 79). In this situation, members 78 and 79 are subjected to the emitter implant in addition to the n+ implant used for doping n-FET 30. Yet another option is to dope regions 78 and 79 solely by means of implant 41.

5

10

15

20

Once the implant sequence 41-43 has been completed, a thermal anneal at 950° C in an N₂ atmosphere for 60 minutes is performed. This anneal is what actually forms the transistor junctions in the silicon substrate. During the anneal, the previously implanted dopant diffuses from the overlying polysilicon members down into the substrate. In other words, each of the SDE polysilicon members functions as a separate diffusion source — introducing impurities into the underlying substrate to form source, drain, emitter and extrinsic base regions of the respective transistors. This is contrary to conventional methods which rely on either direct ion implantation or direct diffusion from a gaseous impurity source. The present method of utilizing doped polysilicon as a diffusion source provides much better control of diffusion rates and junction depth. The actual junctions formed by this method are characterized as being extremely shallow (necessary for very small device structures) and of a very high quality (resulting in a low leakage transistors).

Diffusion from a polysilicon source also provides a novel way to form submicron polysilicon bipolar emitters. As polysilicon emitters provide superior performance in bipolar transistors, this simple doping method is a very valuable processing feature. This is particularly so in the case in a BiCMOS process where submicron polysilicon emitters and very shallow FET junctions are formed simultaneously. Simultaneous diffusion of n-type and p-type dopants into the substrate to form device junctions saves a considerable number of processing steps and obviates the need to subject the wafer to additional heat cycles.

30 Although the preferred embodiment described above relies on implanting the SDE polysilicon which resides above the source/drain and

emitter regions using separate masks, it is to be understood that alternative ways of doping the polysilicon members are possible. For instance, the SDE polysilicon may be doped using standard diffusion techniques. In any case, alternative doping techniques are considered to be within the spirit and scope of the present invention.

5

15

30

Figures 15A & 15B illustrate a cross section of the wafer following the drive-in (i.e., anneal) of the source/drain, base and emitter regions of FETs 30 and 40 and BJT 20. In detailed summary, source/drain regions 73 and 74 of p-channel FET 40 are formed by diffusion of boron from doped polysilicon members 68 and 69; source/drain regions 75 and 76 of n-channel FET 30 are formed by diffusion of phosphorus from polysilicon members 65 and 66; extrinsic base region 81 is formed by diffusion of boron from doped polysilicon base member 35; and emitter region 77 is formed by diffusion of arsenic from doped polysilicon member 67. In the currently preferred embodiment, the junction depth of regions 73-76 and 81 are on the order approximately 0.17 microns while the junction depth of emitter region 77 is approximately 0.1 micron. Incidentally, the width of emitter region 77 is approximately 0.2 microns in the preferred embodiment of the present invention.

Thus, the single anneal performs two functions: First it distributes the impurities uniformly throughout the polysilicon members. Second, it diffuses those impurities into the silicon substrate to form hypershallow electrical functions. (Note that in Figures 15A & 15B SDE polysilicon member 84 is shown having an associated n-type diffused region 82 directly below it in the silicon substrate. Diffusion region 82 extends down to collector plug 25 to form a low resistance contact with buried layer 13. Although not discussed above, SDE polysilicon member 84 is doped n-type using either emitter implant 43 or n-channel FET implant 41).

The invented BiCMOS process thus far described provides several unique advantages over prior art processes. To begin with, the present invention implements a novel and useful planarization scheme by

controlling the spaces etched into a first layer of polysilicon (i.e., waffelization), across the entire surface of the wafer. Once the waffelized spaces have been etched, planarization is achieved by a second deposition of polysilicon (isotropic) followed by an unmasked anisotropic etch. This produces an extremely high degree of planarization which permits extremely small device structures and spacings to be formed.

A second way that the invented BiCMOS process achieves narrow device spacing is by the method of using a thin field oxide to overcome the problem of extended birds beak. The high coupling capacitance -- which usually results from the use of thin field oxide -- is overcome by the novel metal interconnect scheme which will be described shortly.

Planarization is also aided by the use of amorphous polysilicon, which is naturally very smooth and devoid of any grain structure. These properties facilitate reactive ion etching, making it easier to etch accurate features. The use of amorphous polysilicon layer -- lacking any grain structure -- also avoids creating nonuniformities or asperties in the underlying crystal. This is of particular importance in the situation where the polysilicon layer is etched down to the emitter regions of the bipolar junction transistors.

Another useful addendum to the invented method of fabrication is the ability to add well compensation implants into the process in a self-aligned manner at the same time that lightly doped drain and bipolar base implants are performed. As has been described, no additional masking steps are needed for the well compensation implants. The effect of this implant however will be to reduce the FET junction capacitance by a factor of about three.

Additionally, the presently invented BiCMOS process provides the availability of n-type and p-type FETs in enhancement, depletion or zero threshold types using only a single mask step and implant for each type of FET (i.e., n-type or p-type). This provides great flexibility in circuit design while using a minimal number of mask steps.

10

Finally, a novel and useful way to form polysilicon source/drain structures and a novel way to form deep sub-micron polysilicon bipolar emitters has been described. SDE polysilicon is implanted with appropriate doping types and subsequently acts as a diffusion source to produce hypershallow junctions. Source, drain, extrinsic base and emitter regions are formed simultaneously using a single anneal step. The resulting junction depths are highly controlled and provide superior transistor performance. In addition, SDE polysilicon may also used as a low resistance interconnect layer.

Overall, the combination of each of the above-described novel features in the invented BiCMOS process results in device structures and densities which are superior to those produced by prior art processes. At this stage, any of several methods may be employed to complete the circuit structure. However, as will be discussed in detail, the present invention includes additional novel features which also contribute to enhanced device performance, reliability and layout density.

FORMATION OF THE DEVICE CONTACTS

Referring now to Figures 16A & 16B, a poly 1 gate contact mask is

20 utilized to create opening 86 over gate member 33 of p-channel FET 40.

This mask provides one means of access to the poly 1 (SDE poly, gate members, base contacts, poly interconnects and dummy polysilicon) level of the device structure. The approximate dimension of opening 86 is

0.5 microns in the preferred embodiment.

A salient feature of the invented process is the capability of having gate contacts overlap isolation or oxide regions without compromising the electrical integrity of the circuit. The only requirement with respect to gate contacts is that the gate contact opening 86 must not overlap any SDE polysilicon regions (e.g., 68, 69, etc.) since the result would be an unwanted electrical connection. The capacity to be able to provide a gate contact to poly 1 which can overlap oxide and isolation regions is extremely desirable.

10

15

25

Overlapping gate contacts in this manner greatly improves the overall density of integrated circuit layouts. (Note that in Figures 16A & 16B the dimension of opening 86 is not shown to scale for illustrative purposes.)

After gate contact regions 86 have been opened, the water is subjected to an HF dip to clear any remaining oxide from the tops of the SDE polysilicon members (e.g., members 65-69). Cobalt is then evaporated or sputtered over the entire surface of the water to a thickness of approximately 1000Å.

5

15

20

30

The choice of cobalt is central to the particular metalization scheme employed in this process. Once cobalt silicide is formed on the underlying polysilicon members it becomes impervious to nearly all etchants. The physical properties of cobalt silicide make it unusual among metal silicides since it forms no volatile halides -- meaning, of course, that reactive ion or plasma etchers will not attack or destroy it. Cobalt silicide is also capable of withstanding very high temperatures. This means that the poly 1 gate contacts do not need to be covered with a second polysilicon layer (poly 2) to protect them from poly 2 etch damage later in the process. Hence, an appropriate choice of silicide (cobalt silicide in this case) leads to a significant density improvement. This improvement is especially pronounced in the case of memory cells.

After the cobalt silicide has been formed using a rapid thermal anneal, the unreacted cobalt, i.e., over oxide regions 57 and isolation regions 60 and 61, is stripped away. Recognize that the cobalt silicide lowers the sheet resistance of polysilicon interconnection members (such as member 79) and provides a low contact resistance to the other polysilicon members (e.g., 65-69). This has the salutary effect of spreading the current density up into the cobalt layer and away from the substrate junctions.

A hydrogen alloy at 400° C in forming gas (15% H_2 + 85% N_2) is performed next to introduce hydrogen (H_2) into the isolation regions of the circuit. Hydrogen is intentionally introduced to reduce the surface state charge density, O_{55} . Normally, in prior art processes, this step is done very

late in the process. However, because all of the regions defined up to this point will soon be encased in a nitride layer (which is impenetrable to H_2) the isolation oxide regions are loaded with hydrogen at this point in the process.

Once the cobalt silicide layer regions have been defined, the devices may be tested parametrically to verify their functionality. Following parametric testing, the entire surface of the water is encased in a layer of silicon nitride to a thickness of approximately 1000Å. With reference to Figures 18A & 18B, this layer of silicon nitride is shown as layer 89.

Next, a poly 1/poly 2 contact mask is used to open holes in the nitride layer 89 where poly 2 will contact poly 1. Figures 18A & 18B show openings to source 65 and drain 66 of n-channel FET 30, gate 33 of p-channel FET 40, and poly 1 interconnect members 78 and 79. Obviously, these openings are by way of example and other connections or openings may be made to other poly 1 regions depending on the specific circuit design or layout.

Another basic feature of the presently invented CMOS process is that the contact openings in nitride layer 89 are largely self-aligned. In other words, the contact openings can actually overlap into other regions (e.g., gates, isolation, oxide, etc.) since the reactive ion etch used to etch nitride will not ettack oxide or cobalt silicide. It should be recognized that had another type of silicide been used -- other than cobalt silicide -- it could have been removed during reactive ion etching of nitride layer 89.

25 FORMATION OF SECOND LAYER POLYSILICON RESISTORS AND INTERCONNECTS

Following etching of nitride layer 89, a layer of titanium is patterned on the surface of the water using the well-known patterning technique of metal lift-off. Figures 19A & 19B show the patterned titanium 90 which, in the preferred embodiment, is deposited to a thickness of approximately 1000Å. Titanium layer 90 is used for low resistance poly 2 interconnect tayers and also for low resistance contacts at the contact regions, or

10

20

endpoint terminals, of poly 2 resistors.

10

15

20

25

30

Referring now to Figures 20A & 20B, a second polysilicon layer 93 is deposited over the water following the patterning of titanium layer 90. The deposition of polysilicon layer 93 occurs in a furnace at 580° C to a total thickness of approximately 2000Å. As before, polysilicon layer 93 is amorphous in structure in order to take advantage of the superior planar properties that amorphous polysilicon possesses. A 1000Å silicon nitride layer 94 is then deposited over top of layer 93 using LPCVD techniques. A rapid thermal anneal at 1000° C forms titanium disilicide (TiSi₂) with those regions of polysilicon layer 93 which contact titanium layer 90. Nitride layer 94 protects polysilicon layer 93 from possible contamination during exposure to subsequent processing steps.

Once nitride layer 94 has been deposited a poly 2 mask is aligned to the poly 1/poly 2 contact regions. The poly 2 mask is employed to define poly 2 interconnection layers as well as poly 2 resistors. Following patterning of the poly 2 mask, reactive ion etch is used to vertically etch down through nitride layer 94, poly 2 layer 93, titanium disilicide layer 90 and any underlying silicon nitride present in layer 89. The reactive ion etch however will not etch through cobalt silicide or exposed oxide. This allows a considerable margin for misalignment in the poly 2 mask operation.

Figures 21A & 21B illustrate a cross-section of the wafer following the reactive ion etched used to etch poly 2 and the surrounding nitride layers. As is clearly seen, this etching step defines poly 2 resistor 95 and poly 2 interconnect members 98 and 99. Recall that where poly 2 is used as an interconnect (as in member 98) the bottom portion of the poly 2 layer is completely silicided to form a low resistance interconnection medium. Conversely, where poly 2 is used as a resistive element (such as member 95) only the terminals or contact regions are silicided. Accordingly, resistor 95 connects gate 33 with source polysilicon contact member 65. In its undoped state, polysilicon has an extremely high resistivity -- on the order of 109 ohms/square -- versus about 1 ohm/square for completely

silicided polysilicon.

10

15

20

25

30

A significant by-product of the above-described method of defining poly 2 regions, is that it makes it possible to create titanium silicide lines which are much narrower than would normally be achievable using ordinary lithographic techniques (e.g., 0.1 micron interconnection lines). The reason for this is that the titenium interconnection lines in this process are first patterned and defined using a first mask layer (i.e., titanium metal-lift-off mask), and then subsequently etched using a second mask (i.e., poly 2 mask). Taking advantage of the fact that mask alignment is considerably more accurate than the optical resolution for a given mask layer, these two masking steps (along with the vertical reactive ion etch) can be used to define features having extremely narrow dimensions. In other words, this method provides a way of generating sub-micron, or sub-lithographic features by taking the image defined by one mask and moving it to two masks. Therefore, although Figures 21A & 21B show titanium silicide layers having dimensions defined by the lift-off pattern a single mask, it is appreciated that much narrower interconnection lines - on the order of tenths of microns -- may also be achieved.

After the poly 2 layer has been etched, another nitride layer is then deposited over the entire water to a thickness of approximately 1000Å. In Figures 22A & 22B this nitride layer is represented by layer 101. Silicon nitride layer 101 is used to envelop and protect all of the devices.

Although at this point it appears that our goal of complete planarization has been defeated by the poly 2 deposition. It should be emphasized that the design rules in the preferred embodiment process prohibit direct contacts to polysilicon 2. Consequently, planarization actually remains intact since all contacts still reside in the same general plane. Moreover, as will be seen, additional steps are taken in the invented process to assure that planarization is re-established at every level of the metalization process.

5

10

20

FORMATION OF THE AIR-BRIDGE METAL INTERCONNECTS

In the remaining steps, a process to form a planarized, air-bridge interconnect system will be described. As a brief introduction, in this process an organic material (e.g., polyimide) is used as a temporary interlayer dielectric. Pedestal contacts and pedestal vias are utilized along with etch back planarization techniques to achieve a highly planar multilayer metalization scheme. In the currently preferred embodiment this metalization scheme is implemented with two layers of metal. It should be understood, however, that the process is not strictly limited in the number of layers which may be implemented.

Referring to Figures 22A & 22B, a cross-section of the water is illustrated following the definition of the pedestal contact openings 97. Pedestal contact openings 97 are formed using conventional masking techniques to etch nitride layer 101 in those areas where pedestal contacts will be placed. For example, pedestal contact openings are positioned over polysilicon member 84, member 67 and isolation region 61.

After the pedestal contacts have been formed, planarization is re-established at the next highest level using a method which is radically different when compared to prior art techniques. In the prior art, oxide openings to the underlying substrate are usually filled with metal. The metal is deposited over all of the irregularities or steps present in the substrate surface. In contrast, pedestal contacts in the invented process protrude above the silicon surface to a height which is above the highest irregularity in the unplanarized surface.

Observe that in Figures 22A & 22B, the contact openings that are etched through nitride layer 101 are highly self-aligned. Once again, because the reactive ion etch used to etch silicon nitride is impervious to oxide or cobalt silicide, the contact openings for the pedestals can extend over the bases and gates of devices. They can also extend over isolation regions. This self-aligning contact scheme, by itself, results in a fundamental

improvement in overall circuit density.

Contacts must normally be kept well away from gates and field exide regions to prevent either a gate or a field short. Generally, this rule results in a minimum transistor which consists of a tiny active region with the adjacent contacts being disposed in huge open areas around the periphery of the transistor. Obviously this wastes a great deal of silicon area. Conversely, in the invented BiCMOS process contacts can have more than 50% of their total area residing over gate or field exide regions.

Referring now to Figures 23A & 23B, a metal lift-off technique is used to define the pedestal contacts. Photoresist is first deposited over the water and then developed, exposing pedestal contact openings down to the substrate. A titanium layer 105 is then deposited to a thickness of 1000Å. This is followed by a platinum deposition to form layer 106 to a thickness of 1000Å. A gold layer 102 is then deposited to a thickness somewhere between 7000-10,000Å.

Gold is the main current carrying member of the pedestal while platinum provides a barrier to prevent gold diffusion into the underlying substrate region. Titanium is included in the structure because of its low resistivity and its propensity to adhere to silicon. Because gold has one of the highest electromigration resistance levels of all the metals, this pedestal contact structure -- comprising titanium, platinum and gold -- is one of the most reliable metal contacts that may be produced in semiconductor technology today.

It is appreciated that because of the relative height of each of the pedestals (9000-12,000Å), more than one layer of photoresist may need to be applied during the lift-off process. To make a thicker pedestal, a thicker photoresist layer is simply applied. Note also that in Figures 23A & 23B, all of the pedestals shown are illustrated as having the same dimension; that is, equal height and width. Although the pedestals are formed to the same height in this process, there is no requirement that any single pedestal be identical to any other pedestal in length and width. After the pedestals have

15

20

been defined, a 410° C alloy in forming gas is performed.

Once alloyed, a layer of polyimide is spun onto the wafer and cured at a temperature of 250° C. Well understood, polyimide is a plastic material with great strength and a very high melting point. The surface tension created from spinning the wafer distributes the polyimide uniformly across the wafer surface. Curing it at 250° C evaporates out the solvent components of the polyimide and forms cross-linked bonds. Bond formation is frequently referred to as "imidization". While curing at 250° C hardens the polyimide, it does not fully imidize the polyimide, as this would make the material virtually impossible to etch.

After curing, the flat polyimide layer has a thickness of approximately 25,000Å. This is sufficiently thick so that the underlying topology has practically no effect on the surface tension of the polyimide. This insures a flat, planar upper surface. A reactive ion etch in 0₂ is then performed to anisotropically etch back the cured polyimide until the tops of the pedestals 102 appear.

When the tops of the pedestal contacts have been exposed, two layers of photoresist are applied over the surface of the wafer. This photoresist film is then masked to define metal 1 interconnection patterns.

20 Metal 1 deposition then takes place. As with the pedestal contacts, a thin layer of titanium is first deposited. This titanium layer is shown in Figures 24A & 24B as layer 107. Next, approximately 10,000Å of gold is deposited as the main current carrier of the metal 1 interconnects 108. Incidentally, the thickness of the titanium layer is approximately 100Å while the gold may range anywhere from between 7000-10,000Å thick in the preferred embodiment. The metal 1 lines 108 are once again defined using metal lift-off. (Note that in Figures 24A & 24B, the polyimide layer is not explicitly shown. Nevertheless, it is understood to be co-planar with the bottom of titanium layer 107.)

Consider that in Figures 24A & 24B, pedestal contact 102 are used in several different capacities. First of all, pedestal contacts 102 provide

10

electrical connection to the underlying transistor structures. For example, emitter 77 of NPN bipolar transistor 20 is coupled directly to metal 1 using a pedestal contact 102. Similarly, the collector contact of BJT 20 is also connected to metal 1 through a pedestal contact.

Pedestals may also function as structural or physical support members. Figures 24A & 24B show a pedestal contact positioned over isolation region 61. The sole function of this pedestal contact is to provide structural support to the overhead metal 1 interconnect lines. Support members are necessary for those applications in which the underlying polyimide is subsequently removed (for reasons to be discussed later). In a sense, the entire metal 1/metal 2 interconnection system is analogous to a kind of microscopic electronic freeway system that carries current in metal lines suspended above the substrate surface.

The particular need to remove the underlying polyimide for certain applications gives rise to various design rule constraints. For instance, in the preferred embodiment of the present invention, the metal 1 interconnect lines have a minimum width of 0.5 microns and a maximum width of 2.0 microns. The maximum width constraint is necessary to allow the polyimide etchant to penetrate to the underlying polyimide. If metal 1 interconnections had no width constraints, it is possible that polyimide underlying a wide area of metal 1 might not be effectively removed. The maximum metal 1 design rule constraint therefore necessitates the use of parallel strips of metal for high power lines.

The design rules in the currently preferred embodiment also restrict the maximum pedestal spacing to approximately 10 microns. Again, this rule is developed from the need to physically support the metal 1 interconnection lines. Pedestal contacts spaced greater than 10 microns apart create a risk of having metal 1 interconnect lines collapse onto the devices below.

30 Referring now to Figures 25A & 25B, via pedestals 110 are shown formed on top of metal 1 lines 10B. These via pedestals are formed using

5

10

15

20

the same processing steps which were used to construct both the pedestal contacts 102 and metal interconnect lines 108. That is, the polyimide layer is first spun and cured onto the wafer, and then vertically etched exposing only the tops of the metal 1 interconnect lines. Then, using a two-layer photoresist metal lift-off technique, a titanium layer 109 is first deposited followed by a gold deposition to form the via pedestals 110. The thicknesses of layers 109 and 110 are approximately 100Å and 7000-10,000Å, respectively. It should be understood that in each of the layers 105, 107 and 109, chrome or some other similar metal, may be substituted for titanium — the purpose of these layers being only to provide improved adhesion for the overlying gold.

After the lift-off process which defines via pedestals 110, another layer of polyimide is spun on the wafer, cured and then vertically etched back to expose the tops of via pedestals 110. Metal 2 interconnections are then formed in the same manner as has been described in connection with the metal 1 interconnection lines.

Figures 26A & 26B show the completed metal interconnection system with titanium adhesion layer 111 directly below metal 2 interconnect lines 112. One of the beneficial features of the presently invented interconnect structure is that vias can sit directly over device contacts. This is highly desirable for realizing high density circuits and is generally prohibited in conventional processes.

Although it is not shown in any of the Figures, it is entirely possible to have a metal interconnection stack consisting of a pedestal contact 102, a metal 1 layer 108 and a via pedestal 110 whose sole function is to provide physical and structural support for the metal 2 interconnection fines 112. Such a metal stack would have its pedestal contact formed over oxide or isolation regions and would not provide any electrical connection to the underlying layers (e.g., substrate).

To form bonding pads, a polyimide is spun and cured and subsequently etched back to the tops of the metal 2 interconnect lines 112.

10

15

20

25

Titanium (approximately 100Å) followed by gold (approximately 1000Å) is then deposited across the entire surface of the wafer to provide a plating base for the bonding pads. Next, the wafer is photolithographically masked to open the bonding pad locations. An electroplate of gold 25 microns thick is then deposited. This thickness of gold is preferred to accommodate tape automated bonding (TAB) following lift-off of the thick gold layer.

Although this completes the integrated circuit, it remains highly capacitive since polyimide is still present between the metal 2/metal 1/silicon surface regions. For those applications which are not particularly sensitive to this type of capacitance, or are relatively immune to the hydroscopic properties of polyimide, another layer of protective polyimide could be spun and cured over the surface of the wafer for mechanical protection. The wafer would then be sawn, broken, and packaged into a suitable plastic package. This would represent a low-cost, low-performance type of device.

For higher performance devices, the water is placed back into a reactive ion etch (0₂) until all of the polyimide not protected by metal (in a vertical direction) is etched out. Recall that a reactive ion etch is highly directional, i.e., anisotropic. Etching the polyimide in this way would allow opening of underlying polysilicon fuses. These fuses may be opened or blown energetically (e.g., laser) as required for the particular application. For example, redundant columns and memory cells often rely on the use of polysilicon fuses for interconnection.

Of course, this requires careful mask design of the metal 1 and metal 2 mask layers so that polysilicon fuse locations are never covered by metal 1 or metal 2 interconnect. Once the fuses have been opened, another layer of polyimide may be spun and cured over the water for mechanical protection. The water may then be backlapped and sawed into individual die. At this stage, the die could be bonded into any type of appropriate package (e.g., DIP, leadless chip carrier, etc.) or directly into a printed circuit board.

10

15

25

Once bonded, an oxygen asher etch (highly isotropic) is used to clear all of the polyimide from the die surface. The oxygen plasma ashers the polyimide (i.e., converts it to gas) out from under the metal interconnect lines leaving them supported solely by their pedestals. This means that, each metal interconnect line is now completely suspended in space. This results in the lowest possible parasitic capacitance for improved device speed at lower power levels.

As an alternative, an anisotropic etch may be used to remove the polyimide between the lines but not directly under them. This would result in a compromise structure wherein lateral coupling capacitance is reduced, but vertical coupling capacitance is not. Once the underlying polyimide is removed the part may be baked and encapsulated with a lid.

Where air bridge interconnects are undesirable or unnecessary, another alternative exists. In this approach, after the polyimide has been removed the die or wafer is coated with colloidal suspension of Teflon in water. This substance is then died and sintered to fuse the Teflon into a solid mass. The Teflon plus water mixture flows by capillary action into the spaces or voids left by the ashering process. After sintering it will appear as if it has simply replaced the polyimide between the metal lines. Because Teflon has one of the lower dielectric constants of any solid, the resulting structure will have minimum parasitic capacitance for a solid structure. This technique can also be performed using plasma techniques to deposit an appropriate fluorocarbon polymer. Thus, other materials or polymers may be implemented using appropriate deposition processes.

25

30

10

APPLICATIONS OF THE PRESENTLY INVENTED BICMOS PROCESS

The use of an air-bridge interconnection system also opens up the possibility of many new and exciting applications in the semiconductor technology field. For instance, instead of backfilling the air spaces with Teflon or polyimide, an organic material such as a liquid crystal or a Kerr cell

5

15

20

25

material (e.g., nitrobenzene) may be substituted. By talloring the air bridges to precise dimensions and backfilling with an organic material such as nitrobenzene, optical waveguide modulators may be created. Optical waveguide modulators are exceedingly useful for certain communications applications.

Careful selection of electrical/optical materials can also produce phase shifts in various signals to achieve intensity modulation. This would make it possible to build, by way of example, a 10,000 channel gigahertz speed optical modulator. Electro-optical semiconductor technology could result in devices wherein portions of a semiconductor circuit furnish intelligence while other sections provide modulation or optical coupling to laser-based interconnections. Accordingly, the air bridge interconnection scheme of the presently invented BiCMOS process is extremely rich and diverse in its possibilities.

In addition to organic materials, chemical or even biological materials may also be backfilled or implanted into the air bridge interconnect gaps to produce a heretofore unknown type of semiconductor device. For instance, a biological material or tissue could be backfilled into the air bridge voids as part of a biomedical device to be implanted into a biological organism. An appropriate filler material could provide the proper interface with the host organism, thereby eliminating or reducing the possibility of rejection. It is conceivable that such devices could find future application in heart pace-makers, kidney implants, brain implants, and so on.

Thus, a process for simultaneously fabricating MOS-type semiconductor devices and bipolar junction transistors on the same silicon substrate, has been described.

The present application has been divided out of U.K. patent Application No. 9213519.3 (2256527), [derived from PCT/US91/00211] in which there is described and claimed a method of fabricating MOSFETs and BJTs in a substrate using a process of patterning a first conductive layer with equal spacings and providing a second conductive layer in the spacings.

CLAINS

- 1. In a process for fabricating an integrated circuit (IC) which includes a plurality of devices coupled together by a system of metal interconnects, a method of forming said system above the surface of a semiconductor substrate comprising the steps of:
- (a) forming a plurality of conductive pedestals on said surface of said substrate, a first portion of said pedestals forming electrical contacts to said devices, said pedestals being formed to a predetermined height such that the top surface of said pedestals is higher than any feature of said substrate;
- (b) depositing a first layer of polyimide over said substrate to a thickness which is sufficient to cover said pedestals;
- (c) anisotropically etching said first polyimide layer until the upper surface of said first polyimide layer is roughly coplanar with said top surface of said pedestals; and
- (d) forming a first set of metal interconnect lines over said first polyimide layer, thereby electrically coupling selected ones of said pedestals.
- 2. The method as defined in claim 1, further comprising the steps of:
- (e) forming via contacts on said metal interconnect lines;
- (f) depositing a second layer of polyimide to completely cover said metal interconnect lines;
- (g) anisotropically etching said second polyimide layer until the upper surface of said second polyimide layer is roughly coplanar with the top surface of said metal lines;

- (h) forming a second set of metal interconnect lines over said second polyimide layer, thereby electrically coupling selected ones of said via contacts.
- 3. The method as defined in claim 2, including the additional step of removing said first and second polyimide layers to reduce parasitic coupling capacitance between said first and second set of interconnect lines and said substrate.
- 4. The method as defined in claim 3, wherein a second portion of said pedestals solely function as physical support members for said first set of interconnect lines.
- 5. The method as defined in claim 3, wherein said pedestal contacts, said metal interconnect lines, said via contacts and said second metal interconnect lines are formed using a metal lift-off technique.
- 6. The method as defined in claim 4, wherein step (a) comprises the steps of:

masking said substrate to define the lateral extent of each of said pedestals, said lateral extent being permitted to partially overlapping the base, date and/or isolation regions of said devices;

depositing a layer of titanium; depositing a layer of platinum; depositing a layer of gold.

7. The method as defined in claim 6, wherein said titanium, platinum and gold layers are approximately 1000Å, 1000Å and 7000-10,000Å thick, respectively.

- 8. The method as defined in claim 4, wherein said first and second set of metal interconnect lines each comprise a layer of gold deposited over a layer of titanium.
- 9. The method as defined in claim 8, wherein said titanium layer is approximately 1000Å thick and said gold layer is approximately 7000-10,000Å thick.
- 10. The method as defined in claim 9, including the additional step of removing said first and second polyimide layers to reduce parasitic coupling capacitance between said first and second set of interconnect lines and said substrate.
- 11. The method as defined in claim 2, wherein certain of said via contacts are formed over the device contacts of said IC.
- 12. The method as defined in claim 4, further comprising the step of backfilling the areas where said first polyimide layer has been removed with a fluorocarbon polymer acting as a dielectric.
- 13. The method as defined in claim 12, wherein said fluorocarbon polymer comprises $TEFLON^{TM}$ material.
- 14. The method as defined in claim 3, further comprising the step of backfilling the areas where said first and second polyimide layers has been removed with a fluorocarbon polymer acting as a dielectric.
- 15. The method as defined in claim 14, wherein said fluorocarbon polymer comprises Teflon.

- 16. The method as defined in claim 3, further comprising the step of forming electro-optical modulators in at least a portion of the areas where said first polyimide layer has been removed by filling said areas with an organic material.
- 17. The method as defined in claim 16, wherein said organic material comprises nitrobenzene.
- 18. The method as defined in claim 3, further comprising the step of backfilling at least a portion of the areas where said first and second polyimide layers have been removed with a material suitable for electrically coupling said IC to biological tissue within an organism.
- 19. In a process for fabricating an integrated circuit (IC) comprising a plurality of semiconductor devices formed in a silicon substrate, a method of forming planarized, air-bridge interconnects comprising the steps of:
- (a) forming pedestal contacts raised above from the surface of said substrate, said pedestal contacts having a thickness which allows the tops of said pedestal contacts to be either substantially co-planar with or higher than the highest feature of said IC, at least one of said pedestal contacts being electrically coupled to said devices;
- (b) depositing a first layer of polyimide over said IC to a thickness sufficient to cover said tops of said pedestal contacts;
- (c) vertically etching said first polyimide layer until the top surface of said first polyimide layer is approximately co-planar with said tops of said

pedestal contacts.

- (d) forming first metal interconnect lines electrically coupled to said pedestal contacts over said first polyimide layer;
- (e) forming via contacts on said first metal
 interconnect lines;
- (f) depositing a second layer of polyimide over said IC to a thickness sufficient to cover said via contacts;
- (g) vertically etching said second polyimide layer until the top surface of said second polyimide layer is approximately co-planar with the tops of said via contacts;
- (h) forming second metal interconnect lines electrically coupled to said via contacts over said second polyimide layer.
- 20. The method as defined in claim 19, further comprising the step of ashering said first and second polyimide layers under said first and said second metal interconnect lines to leave voids thereunder, said first and second lines being supported only by said pedestal contacts and said via contacts.
- 21. The method as defined in claim 20, further comprising the additional steps of:

coating said IC with a colloidal suspension of Teflon to fill said voids;

fusing said Teflon into a solid mass to form an interlayer dielectric between said first and second metal interconnect lines and said substrate.

22. The method as defined in claim 21, wherein said Teflon covers said second metal interconnect lines to

provide mechanical strength and contaminant protection for said IC.

Patents Act 1977 -47 - Examiner's report to the Comptroller under Section 17 (The Search report)	Application number GB 9320218.2	
Relevant Technical Fields (i) UK Cl (Ed.L) H1K (KGF, KHAB, KHAX)	Search Examiner W A MORRIS	
(ii) Int Cl (Ed.5) H01L	Date of completion of Search 16 NOVEMBER 93	
Databases (see below) (i) UK Patent Office collections of GB, EP, WO and US patent specifications.	Documents considered relevant following a search in respect of Claims:- 1-22	

Categories of documents

X:	Document indicating lack of novelty or of inventive step.	P:	Document published on or after the declared priority date

- Y: Document indicating lack of inventive step if combined with one or more other documents of the same category.

 E: Patent document published on or after, but with priority date earlier than, the filing date of the present application.
- A: Document indicating technological background and/or state of the art.

 &: Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages	Relevant to claim(s)
	NONE	
		l.
		•
	·	
		·
	·	

Databases: The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).